



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ
ΥΠΟΛΟΓΙΣΤΩΝ

Τομέας Τεχνολογίας Πληροφορικής και Υπολογιστών

Υλοποίηση Μοντέλων Βαθιάς Μάθησης και
Τεχνικών Μεταφοράς Γνώσης σε Ιατρικές Εικόνες

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

ΦΙΛΙΠΠΟΣ
ΣΚΟΒΕΛΕΦ
ΟΡΦΑΝΟΥΔΑΚΗΣ

Επιβλέπων: Στέφανος Κόλλιας
Καθηγητής Ε.Μ.Π.

Συνεπιβλέπουσα: Παρασκευή Τζούβελη
ΕΔΙΠ

Αθήνα, Οκτώβριος 2020



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ
ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
Τομέας Τεχνολογίας Πληροφορικής και Υπολογιστών

Υλοποίηση Μοντέλων Βαθιάς Μάθησης και
Τεχνικών Μεταφοράς Γνώσης σε Ιατρικές Εικόνες

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

ΦΙΛΙΠΠΟΣ
ΣΚΟΒΕΛΕΦ
ΟΡΦΑΝΟΥΔΑΚΗΣ

Επιβλέπων: Στέφανος Κόλλιας
Καθηγητής Ε.Μ.Π.

Συνεπιβλέπουσα: Παρασκευή Τζούβελη
ΕΔΙΠ

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την

.....
Στέφανος Κόλλιας
Καθηγητής Ε.Μ.Π.

.....
Γεώργιος Στάμου
Αν. Καθηγητής Ε.Μ.Π

.....
Ανδρέας - Γεώργιος
Σταψλοπάτης
Καθηγητής Ε.Μ.Π

Αθήνα, Οκτώβριος 2020.

.....
Φίλιππος Σκόβελεφ Ορφανουδάκης
Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών
Ε.Μ.Π.Ι.

Copyright © Φίλιππος Σκόβελεφ Ορφανουδάκης, 2020
Με επιφύλαξη παντός δικαιώματος. All rights reserved

Απαγορεύεται η αντιγραφή, απουθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, απουθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα. Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

Περίληψη

Στο πλαίσιο αυτής της διπλωματικής εργασίας, μελετήθηκε η συμπεριφορά συστημάτων βαθιάς μάθησης με στόχο την ταξινόμηση ακτινογραφιών πνευμόνων ανθρώπων σε 3 κατηγορίες-καταστάσεις: πνευμονία, covid-19, κανονικές.

Για αυτόν τον σκοπό, χρησιμοποιήθηκε μια συλλογή δεδομένων με 18356 ακτινογραφίες για την ανάπτυξη διαφορετικών μοντέλων αλλά και προσεγγίσεων στο πρόβλημα της ταξινόμησης. Η βέλτιστη τελική ακρίβεια που επιτεύχθηκε είναι 93% και εξήχθησαν χρήσιμες πληροφορίες για την επιλογή των υπερπαραμέτρων.

Πρώτο βήμα ήταν η εξισορρόπηση της συλλογής που συλλέχθηκε καθώς οι ακτινογραφίες της κατηγορίας covid-19 ήταν αισθητά λιγότερες. Για την επίτευξη αυτού του στόχου αναπτύχθηκε ένα Παραγωγικό Αντιπαλικό Δίκτυο (Generative Adversarial Network), όπως επίσης αξιοποιήθηκε η μέθοδος της Επαύξησης Δεδομένων (Data Augmentation).

Στη συνέχεια, χρησιμοποιήθηκαν τρία προεκπαιδευμένα μοντέλα το VGG19, το ResNet152V2 και το CheXNet/Dense121 για να εφαρμοστούν μέθοδοι μεταφοράς γνώσης (Transfer Learning).

Για την τελική επιλογή των υπερπαραμέτρων αξιοποιήθηκε ο αλγόριθμος Βελτιστοποίησης Συμήνους Σωματιδίων (Particle Swarm Optimization - PSO), ο οποίος μετά την ολοκλήρωση του ανέδειξε τις βέλτιστες αρχιτεκτονικές που κατάφερε να βρει.

Για κάθε ένα από τα παραπάνω προτεινόμενα μοντέλα πραγματοποιήθηκε η διαδικασία βελτιστοποίησης υπερπαραμέτρων δύο φορές: μια φορά με τη συλλογή δεδομένων που εξισορροπήθηκε και μια φορά με τη συλλογή δεδομένων που δεν εξισορροπήθηκε. Αφού βρέθηκε η τελική αρχιτεκτονική για κάθε περίπτωση, εφαρμόστηκε η συλλογή δεδομένων που διατίθεται ως τεστ και εξήχθησαν τα αποτελέσματα.

Τέλος, συγκρίθηκαν τα αποτελέσματα των μοντέλων με μια αρκετά διαφορετική συλλογή δεδομένων. Μια συλλογή που αποτελείται από λιγότερες ακτινογραφίες με αρκετά καλύτερη ποιότητα. Με αυτόν τον τρόπο βγήκαν τα συμπεράσματα για το ποιο χαρακτηριστικό - ποιότητα ή ποσότητα - είναι πιο χρήσιμο στην περίπτωση μας.

Λέξεις κλειδιά: COVID-19, Ακτινογραφίες, Βαθιά Μάθηση, Μεταφορά Γνώσης, Προεκπαιδευμένα Μοντέλα, Παραγωγικά Αντιπαλικά Δίκτυα, Βελτιστοποίηση Συμήνους Σωματιδίων

Abstract

In the context of this diploma thesis, the behavior of deep learning systems was studied with the aim of classifying human lung radiographs into 3 categories-conditions: pneumonia, covid-19 and normal.

For this purpose, a data collection with 18,356 x-rays was used for the development of different models as well as approaches to deal with this classification problem. An optimum final accuracy of 93% was achieved and useful information was extracted concerning the selection of hyperparameters.

The first step was to balance our collection since the x-rays of the covid19 class were significantly fewer. To achieve this goal, a Generative Adversarial Network was developed and Data Augmentation methods were also utilized.

Then, three pre-trained models, VGG19, ResNet152V2 and CheXNet / Dense121 were used to apply Transfer Learning methods.

For the final selection of the hyperparameters, the Particle Swarm Optimization (PSO) algorithm was used, which after its completion, displayed the optimal structures that it managed to find.

For each of the above proposed models, the hyperparameter optimization process was performed twice: once with the balanced data collection and once with the unbalanced data collection. When the final structure for each case was found, the data collection which was used as a test was applied and the results were exported.

Finally, the results of the models were compared with a substantially different data collection, which consisted of fewer x-rays of much better quality. Conclusions were drawn as to which characteristic- quality or quantity - is most useful in our case.

Keywords: COVID-19, X-rays, Deep Learning, Transfer Learning , Pretrained Models, Generative Adversarial Networks, Particle Swarm Optimization

Ευχαριστίες

Σε αυτό το σημείο θα ήθελα να ευχαριστήσω θερμά τον επιβλέποντα καθηγητή μου κ. Στέφανο Κόλλια για την ευκαιρία και εμπιστοσύνη που μου έδωσε να ασχοληθώ με ένα επίκαιρο και κατά μια έννοια φιλανθρωπικό θέμα.

Επίσης, θα ήθελα να ευχαριστήσω τα μέλη της επιτροπής τον Καθηγητή κ. Ανδρέα-Γιώργο Σταυρολοπάτη και τον Αν. Καθηγητή κ. Γεώργιο Στάμου για την τιμή που μου έκαναν να συμμετάσχουν στην επιτροπή εξέτασης της εργασίας.

Ιδιαίτερες ευχαριστίες θα ήθελα να δώσω στην Διδάκτορα και επιβλέπουσά μου Παφασκευή Τζούβελη η οποία από την πρώτη στιγμή προσέφερε ουσιαστική βοήθεια τόσο σε πρακτικό όσο και ψυχολογικό επίπεδο. Χωρίς την στήριξη, την κατανόηση και τις συμβουλές της δεν θα είχε ολοκληρωθεί η εργασία.

Τέλος θα ήθελα να ευχαριστήσω την οικογένειά μου που ήταν δίπλα μου σε όλη αυτήν την πορεία ακόμα και όταν κυριαρχούσε η πίεση συνέχιζαν να μου δίνουν κίνητρο.

Φίλιππος Σκόβελεφ Ορφανουδάκης
Αθήνα, Οκτώβριος 2020

Περιεχόμενα

1 Εισαγωγή	13
1.1 Δομή Διπλωματικής Εργασίας	14
1.2 Συγγενείς Εργασίες	14
2 Θεωρητικό Υπόβαθρο	16
2.1 Παραγωγικό Αντιπαλικό Δίκτυο	16
2.1.1 Παραγωγικό Δίκτυο	17
2.1.2 Διαχωριστικό Δίκτυο	19
2.1.3 Διαδικασία εκπαίδευσης	20
2.1.4 Διάφορες Χρήσεις Παραγωγικών Αντιπαλικών Δικτύων	23
2.2 Μεταφορά Γνώσης	25
2.2.1 Κατηγορίες Μεταφοράς Γνώσης στο πεδίο της Μηχανικής Μάθησης	26
2.2.2 Τεχνικές και Μέθοδοι Μεταφοράς Γνώσης στο πεδίο της Βασιάς Μάθησης	27
2.2.3 Μεταφορά Γνώσης στο πεδίο της Υπολογιστικής Όρασης	29
2.3 Επαύξηση Δεδομένων	30
2.3.1 Επαύξηση Δεδομένων στην Υπολογιστική Όραση	30
2.3.2 Επαύξηση Δεδομένων και Πλεονεκτήματα	31
2.4 Βελτιστοποίηση Συμήνους Σωματιδίων	32
2.4.1 Εισαγωγή στον Αλγόριθμο	32
2.4.2 Παράμετροι Αλγορίθμου	33
2.4.3 Βελτιστοποίηση Συμήνους Σωματιδίων στα Νευρωνικά Δίκτυα	35
3 Ανάλυση Κατασκευής Μοντέλου	36
3.1 Δεδομένα	36
3.2 Μοντέλο για Ισορρόπηση Δεδομένων	40
3.2.1 Παραγωγικό Αντιπαλικό Δίκτυο	40
3.2.2 Διαδικασία Εκπαίδευσης ΠΑΔ	44
3.2.3 Επαύξηση Δεδομένων	46

3.3	Επιλογή και Επεξεργασία Προεκπαιδευμένων Μοντέλων	48
3.3.1	VGG19	49
3.3.2	ResNet152V2	50
3.3.3	CheXNet/Dense121	52
3.3.4	Ρύθμιση Στρωμάτων και Υπερπαραμέτρων	54
3.3.5	Εφαρμογή Βελτιστοποίησης Συμήνους Σωματιδίων και Μετρικές	56
4	Πειραματικά Δεδομένα	58
4.1	Παραγωγικό Αντιπαλικό Δίκτυο	58
4.1.1	Εικόνες Μεγέθους (128,128,1)	58
4.1.2	Εικόνες Μεγέθους (64,64,3)	61
4.1.3	Εικόνες Μεγέθους (64,64,1)	64
4.1.4	Μοντέλο χωρίς Θόρυβο και Εξομάλυνση Κλάσεων	67
4.1.5	Μοντέλο με μικρότερο αριθμό στρωμάτων	71
4.1.6	Μοντέλο με μεγαλύτερο αριθμό στρωμάτων	74
4.2	Μεταφορά Γνώσης και Εφαρμογή Αλγόριθμου Βελτιστοποίησης Συμήνους Σωματιδίων	78
4.2.1	VGG19	79
4.2.2	ResNet152V2	84
4.2.3	CheXNet/Dense121	90
4.3	Τελικά αποτελέσματα	93
4.4	Σημασία Ποιότητας και Ποσότητας στα Δεδομένα	100
5	Σύνοψη και Προτάσεις	104

Κατάλογος Εικόνων

2.1	Παραγωγή πλασματικών εικόνων μέσω ΠΑΔ [1]	17
2.2	Βελτίωση ποιότητας εικόνας μέσω ΠΑΔ [66]	17
2.3	Δομή Παραγωγικού Δικτύου [2]	19
2.4	Δομή Διαχωριστικού Δικτύου[3]	20
2.5	Έξοδος CycleGAN [7]	25
2.6	Μοντέλο Μεταφοράς Γνώσης [9]	28
2.7	Error rate [11]	29
2.8	Περιστροφή και Οριζόντια Αναστροφή	31
3.1	Κατανομή Δεδομένων	38
3.2	Κατανομή Δεδομένων	39
3.3	Covid-19	39
3.4	Normal	39
3.5	Pneumonia	40
3.6	Δομή Παραγωγικού Δικτύου	41
3.7	Δομή Διαχωριστικού Δικτύου	43
3.8	Αποτελέσματα ΠΑΔ	45
3.9	Παράδειγμα Εισόδου - (64,64,1)	45
3.10	Παράδειγμα Εισόδου - (64,64,1)	45
3.11	Train/Validation Distribution	47
3.12	Τροποποιημένη Εικόνα 1	48
3.13	Τροποποιημένη Εικόνα 2	48
3.14	VGG19 Architecture [30]	50
3.15	Τμήμα Ανατροφοδότησης [36]	51
3.16	ResNet152 architecture[35]	52
3.17	ResNet152 vs ResNet152v2 [20]	52
3.18	DenseBlocks [21]	54
3.19	DenseNets [22]	54
4.1	Διάγραμμα σφάλματος για (128,128,1) εικόνες	59

4.2	200 Εποχές (128,128,1)	59
4.3	500 Εποχές (128,128,1)	60
4.4	800 Εποχές (128,128,1)	60
4.5	1000 Εποχές (128,128,1)	61
4.6	Διάγραμμα σφάλματος για (64,64,3) εικόνες	62
4.7	200 Εποχές (64,64,3)	62
4.8	500 Εποχές (64,64,3)	63
4.9	800 Εποχές (64,64,3)	63
4.10	1000 Εποχές (64,64,3)	64
4.11	Διάγραμμα σφάλματος για (64,64,1) εικόνες	65
4.12	200 Εποχές (64,64,1)	65
4.13	500 Εποχές (64,64,1)	66
4.14	800 Εποχές (64,64,1)	66
4.15	1000 Εποχές (64,64,1)	67
4.16	Διάγραμμα σφάλματος για (64,64,1) εικόνες - Ασταθές	68
4.17	Διάγραμμα ακρίβειας για (64,64,1) εικόνες - Ασταθές	68
4.18	200 Εποχές (64,64,1) - Ασταθές	69
4.19	500 Εποχές (64,64,1) - Ασταθές	69
4.20	800 Εποχές (64,64,1) - Ασταθές	70
4.21	1000 Εποχές (64,64,1) - Ασταθές	70
4.22	Διάγραμμα σφάλματος για (64,64,1) εικόνες - Μικρότερο Μοντέλο	72
4.23	200 Εποχές (64,64,1) - Μικρότερο Μοντέλο	72
4.24	500 Εποχές (64,64,1) - Μικρότερο Μοντέλο	73
4.25	800 Εποχές (64,64,1) - Μικρότερο Μοντέλο	73
4.26	1000 Εποχές (64,64,1) - Μικρότερο Μοντέλο	74
4.27	Διάγραμμα σφάλματος για (64,64,1) εικόνες - Μεγαλύτερο Μοντέλο	75
4.28	200 Εποχές (64,64,1) - Μεγαλύτερο Μοντέλο	76
4.29	500 Εποχές (64,64,1) - Μεγαλύτερο Μοντέλο	76
4.30	800 Εποχές (64,64,1) - Μεγαλύτερο Μοντέλο	77
4.31	1000 Εποχές (64,64,1) - Μεγαλύτερο Μοντέλο	77
4.32	vgg_simple 1 Accuracy	80
4.33	vgg_simple 1 Loss	80
4.34	vgg_simple 2 Accuracy	80
4.35	vgg_simple 2 Loss	80
4.36	vgg_enhanced 1 Accuracy	81
4.37	vgg_enhanced 1 Loss	81
4.38	vgg_enhanced 2 Accuracy	82
4.39	vgg_enhanced 2 Loss	82
4.40	vgg_lstm_simple 1 Accuracy	83
4.41	vgg_lstm_simple 1 Loss	83

4.42 vgg_lstm_simple 2 Accuracy	83
4.43 vgg_lstm_simple 2 Loss	83
4.44 vgg_lstm_enhanced 1 Accuracy	84
4.45 vgg_lstm_enhanced 1 Loss	84
4.46 resnet_simple 1 Accuracy	85
4.47 resnet_simple 1 Loss	85
4.48 resnet_simple 2 Accuracy	85
4.49 resnet_simple 2 Loss	85
4.50 resnet_enhanced 1 Accuracy	86
4.51 resnet_enhanced 1 Loss	86
4.52 resnet_enhanced 2 Accuracy	87
4.53 resnet_enhanced 2 Loss	87
4.54 resnet_lstm_simple 1 Accuracy	88
4.55 resnet_lstm_simple 1 Loss	88
4.56 resnet_lstm_simple 2 Accuracy	88
4.57 resnet_lstm_simple 2 Loss	88
4.58 resnet_lstm_enhanced 1 Accuracy	89
4.59 resnet_lstm_enhanced 1 Loss	89
4.60 resnet_lstm_enhanced 2 Accuracy	89
4.61 resnet_lstm_enhanced 2 Loss	89
4.62 chexnet_simple 1 Accuracy	90
4.63 chexnet_simple 1 Loss	90
4.64 chexnet_enhanced 1 Accuracy	91
4.65 chexnet_enhanced 1 Loss	91
4.66 chexnet_lstm_simple 1 Accuracy	92
4.67 chexnet_lstm_simple 1 Loss	92
4.68 chexnet_lstm_simple 2 Accuracy	92
4.69 chexnet_lstm_simple 2 Loss	92
4.70 chexnet_lstm_enhanced 1 Accuracy	93
4.71 chexnet_lstm_enhanced 1 Loss	93
4.72 Ανάλυση (128,128,3)	101
4.73 Ανάλυση (64,64,3)	101
4.74 Ποσότητα - Ποιότητα	102

Κατάλογος Πινάκων

4.1	vgg_simple	80
4.2	vgg_enhanced	81
4.3	vgg_lstm_simple	82
4.4	vgg_lstm_enhanced	84
4.5	resnet_simple	85
4.6	resnet_enhanced	86
4.7	resnet_lstm_simple	87
4.8	resnet_lstm_enhanced	89
4.9	chexnet_simple	90
4.10	chexnet_enhanced	91
4.11	chexnet_lstm_simple	92
4.12	chexnet_lstm_enhanced	93
4.13	Ακρίβεια στο Τεστ Σετ	95
4.14	vgg_simple 1	95
4.15	vgg_simple 2	96
4.16	vgg_enhanced 1	96
4.17	vgg_lstm_simple 1	97
4.18	vgg_lstm_simple 2	97
4.19	vgg_lstm_enhanced 1	98
4.20	resnet_simple 1	98
4.21	resnet_lstm_enhanced 1	99
4.22	chexnet_enhanced 1	99
4.23	Ακρίβεια στο Τεστ Σετ	102
4.24	vgg_simple 1	103

Κεφάλαιο 1

Εισαγωγή

Τον Δεκέμβριο του 2019, στην πόλη Ουχάν της Κίνας, ξεκίνησε η εξάπλωση του νέου κορονοϊού SARS-CoV-2. Αρχικά διαγνώστηκαν σαν περιπτώσεις πνευμονίας και ύστερα στις 9 Ιανουαρίου 2020 ανακοινώθηκε ότι πρόκειται για νέο στέλεχος κορονοϊού. Ως τις 27 Σεπτεμβρίου 2020 έχουν επιβεβαιωθεί πάνω από 33,000,000 κρούσματα, 1,000,000 θάνατοι και έχουν ανακάμψει πάνω από 24,000,000 άνθρωποι.

Το πιο διαδομένο τεστ ανίχνευσης του ιού SARS-CoV-2 ακολουθεί τη μοριακή μέθοδο RT-PCR. Το συγκεκριμένο τεστ υεωρείται το πιο αξιόπιστο καθώς η ευαισθησία του (λανθασμένα αρνητικά) κυμαίνεται περίπου στο 90% (ανάλογα με την μέθοδο συλλογής του δείγματος). Παρόλα αυτά έχει παρατηρηθεί έλλειψη του αντίστοιχου εξοπλισμού σε πολλές χώρες με αυτό να έχει ως συνέπεια τον μη περιορισμό της εξάπλωσης του ιού. Επίσης λόγω των πολλών δειγμάτων και της εκ φύσεως “μεγάλης” χρονικής διάρκειας εξαγωγής αποτελεσμάτων του τεστ, υπάρχει μεγάλη καθυστέρηση στην διάγνωση. Τέλος, ένας περιορισμός της μοριακής μεθόδου RT-PCR είναι ότι αδυνατεί να εντοπίσει άλλες παρεμφερείς ασθένειες που είναι υπεύθυνες για αρκετά συμπτώματα.

Τα παραπάνω μειονεκτήματα, έχουν κινητοποιήσει τον κλάδο της βαθιάς μάθησης να προτείνει μια διαφορετική προσέγγιση στην διάγνωση του ιού, αφού πολλές φορές στο παρελθόν τεχνικές της βαθιάς μάθησης έχουν εφαρμοστεί με επιτυχία σε προβλήματα όπως κατηγοριοποίηση καρκίνου του δέρματος [28] ή αναγνώριση δυσλειτουργιών του εγκεφάλου [64].

Η επιστημονική κοινότητα έχει ασχοληθεί με το συγκεκριμένο ζήτημα από τους πρώτους μήνες της πανδημίας. Από τις πρώτες ενέργειες ήταν ο σχηματισμός διαφόρων συλλογών δεδομένων καθαρά για χρήση έρευνας. Τα δεδομένα αυτά αφορούν είτε ακτινογραφίες είτε αξονικές είτε συλλογή άρθρων είτε ακόμα και χρονοσειρές εξάπλωσης του ιού σε διάφορες χώρες. Κάθε ένα από αυτά τα δεδομένα εφαρμόζεται σε διαφορετική κλάση της μηχανικής και βαθιάς μάθησης.

Στο πλαίσιο της διπλωματικής εργασίας υπήρξε κινητοποίηση για αξιοποιήση των δεδομένων-εικόνων ακτινογραφιών και για εξέταση της συμπεριφοράς διαφόρων μοντέλων με διαφορετικές δομές, με σκοπό την εύρεση της κατάλληλης αρχιτεκτονικής για την ταξινόμηση των εικόνων σε 3 κατηγορίες: πνευμονία, covid19, κανονικές.

Αποφασίστηκε να χρησιμοποιηθεί η τεχνική της Μεταφοράς Γνώσης (Transfer Learning) και της επανεκπαίδευσης (Fine-Tuning) για τη ρύθμιση των υπερπαραμέτρων, σε δύο συλλογές δεδομένων. Η πρώτη συλλογή αποτελείται από τα διαθέσιμα δεδομένα στο διαδίκτυο, τα οποία είναι - μη ισορροπημένα ως προς την κατηγορία covid-19 - και η δεύτερη συλλογή αποτελείται από τα διαθέσιμα δεδομένα που με μοντέλα (Παραγωγικό Αντιπαλικό Δίκτυο - Επαύξηση δεδομένων) έγινε η προσπάθεια για εξισορρόπηση. Τέλος εξήσυχαν τα αποτελέσματα και δύθηκαν προτάσεις για το ποιο μοντέλο είναι κατάλληλο σε διάφορες περιπτώσεις.

1.1 Δομή Διπλωματικής Εργασίας

Στο 1ο Κεφάλαιο, παρουσιάζεται συνοπτικά το έναυσμα για την εκπόνηση της εργασίας όπως επίσης και ονομαστικά οι τεχνικές και η ροή που ακολουθήσαμε.

Στο 2ο Κεφάλαιο, παρατίθεται μια θεωρητική βάση καθώς και η επεξήγηση των εννοιών που αναφέρονται και χρησιμοποιούνται.

Στο 3ο Κεφάλαιο, αναλύεται η δομή όλων των μοντέλων που χρησιμοποιήθηκαν καθώς και οι διαδικασίες που πραγματοποιήθηκαν.

Στο 4ο Κεφάλαιο, παρατίθενται όλα τα πειραματικά αποτελέσματα που οδήγησαν στα τελικά ευρήματα καθώς επίσης συγχρίνεται ο ρόλος της ποιότητας και της ποσότητας των εικόνων στην συγχειριμένη περίπτωση.

Στο 5ο Κεφάλαιο πραγματοποιείται μια σύνοψη όλης της διαδικασίας και παρουσιάζονται μελλοντικές προτάσεις.

1.2 Συγγενείς Εργασίες

Στην διεθνή βιβλιογραφία, έχει μελετηθεί η ανίχνευση και ο εντοπισμός του ιού SARS-CoV-2 με τη χρήση εικόνων.

Η έλλειψη δεδομένων έχει οδηγήσει στην εφαρμογή της Μεταφοράς Γνώσης (Transfer Learning) με αρκετά προεκπαίδευμένα μοντέλα σαν βάση. Έχει χρησιμοποιηθεί το DenseNet201 [38] για ταξινόμηση σε δύο κλάσεις με ακρίβεια 96,25%, το ResNet50 [29] για ταξινόμηση σε 3 κλάσεις με ακρίβεια 96,23% όπως επίσης έχει χρησιμοποιηθεί και ένα από τα μοντέλα που θα επεξεργαστούμε σε αυτή την εργασία, το VGG19 [65] με ακρίβεια 83%. Το κοινό σε όλα αυτά τα μοντέλα

είναι ότι για την εκπαίδευση τους αλλά και για την δοκιμή τους είχαν λιγότερα δεδομένα από αυτά που διατίθενται τώρα.

Για την ενίσχυση της συλλογής δεδομένων έχουν χρησιμοποιηθεί τεχνικές SMOTE [54] πάνω σε πίνακες χαρακτηριστικών (feature vectors) που έχουν εξαχθεί από τις εικόνες. Η μέθοδος που χρησιμοποιήθηκε σε αυτή την εργασία βασίζεται στην παραγωγή εικόνων μέσω ενός Παραγωγικού Αντιπαλικού Δικτύου για το οποίο έγινε η προσπάθεια να επεκταθεί η ιδέα από το άρθρο του Mohamed Loeuy [50] το οποίο δημοσιεύτηκε τον Απρίλιο του 2020, στο συγκεκριμένο άρθρο εξετάστηκε η απόδοση στο AlexNet, στο GoogleNet και στο ResNet18 με την ακρίβεια να είναι 85.19%, 81.48%, 81.48% αντίστοιχα.

Για την εύρεση των βέλτιστων υπερπαραμέτρων έχει χρησιμοποιηθεί ο αλγόριθμος Grey Wolf Optimizer (GWO) [32], ο οποίος βοήθησε να πετύχει το μοντέλο του 97% ακρίβεια, αξίζει να αναφερθεί πως σε αρκετές περιπτώσεις ο αλγόριθμος PSO που χρησιμοποιήθηκε σε αυτή την εργασία έχει πολύ καλύτερη απόδοση από τον GWO [55].

Στο πλαίσιο της εργασίας και της ανάπτυξης των βαθιών νευρωνικών αρχιτεκτονικών μελετήθηκαν οι επόμενες εργασίες που αφορούν σε εφαρμογή βαθιών νευρωνικών δικτύων για ιατρικές προβλέψεις [46], [62], [63], [40], [67], στην εξαγωγή λανθανουσών μεταβλητών από βαθιά νευρωνικά δίκτυα για διαφανείς προβλέψεις [49], [44], για επανεκπαίδευση και προσαρμογή βαθιών νευρωνικών δικτύων σε διαφορετικά σετ δεδομένων [43], [45], [42], για σύνδεση βαθιών νευρωνικών δικτύων και σημασιολογικών τεχνικών [48], [41], [31], για χρήση τεχνικών προσοχής (attention) σε δίκτυα [25, 57] και για σύνθεση νέων αναπαραστάσεων από υπάρχουσες [47, 49]

Κεφάλαιο 2

Θεωρητικό Υπόβαθρο

2.1 Παραγωγικό Αντιπαλικό Δίκτυο

Το Παραγωγικό Αντιπαλικό Δίκτυο (Generative Adversarial Network) ή αλλίως ΠΑΔ (GAN), προτάθηκε για πρώτη φορά το 2014 [33] από τον Ian Goodfellow και τους συνεργάτες του και είναι ένα μοντέλο μηχανικής μάθησης με στοιχεία βασιάς μάθησης όπως τα συνελικτικά δίκτυα (convolutional neural networks).

Τα ΠΑΔ τοποθετούνται στην κατηγορία της μη επιβλεπόμενης μάθησης και στόχος τους είναι να παράξουν πλασματικά δεδομένα που να ταιριάζουν στο μοτίβο των “αυθεντικών” δεδομένων. Για να το επιτύχουν αυτό προσπαθούν να εκτιμήσουν την κατανομή των στοιχείων σε κάθε ένα από τα δεδομένα πχ για εικόνες, τις τιμές των πινάκων που τις αποτελούν. Για τον σκοπό αυτό η δομή των ΠΑΔ χωρίζεται σε δύο τμήματα:

- Παραγωγικό Δίκτυο (Generator): Μοντέλο που παράγει νέα πλασματικά δεδομένα
- Διαχωριστικό Δίκτυο (Discriminator): Μοντέλο που προσπαθεί να ξεχωρίσει το δεδομένα που έχει παράξει το Παραγωγικό Δίκτυο, από τα “αυθεντικά” δεδομένα που ανήκουν στην αρχική συλλογή.



Figure 2.1: Παραγωγή πλασματικών εικόνων μέσω ΠΑΔ [1]

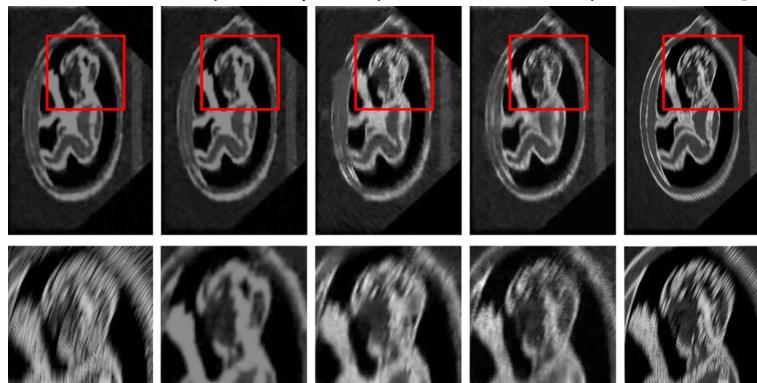


Figure 2.2: Βελτίωση ποιότητας εικόνας μέσω ΠΑΔ [66]

2.1.1 Παραγωγικό Δίκτυο

Το Παραγωγικό Δίκτυο έχει στόχο να παράξει όσο το δυνατόν πιο ρεαλιστικά δεδομένα, δηλαδή να κατανοήσει την κατανομή των δεδομένων εισόδων όπως και το μοτίβο τους.

Πιο συγκεκριμένα, στο κομμάτι της υπολογιστικής όρασης με το οποίο πραγματεύεται η εργασία, τα δεδομένα αφορούν εικόνες, συνεπώς η δομή του Παραγωγικού Δικτύου πρέπει να είναι τέτοια ώστε το τελικό αποτέλεσμα που θα εξάγει, να είναι μια εικόνα ή αλλιώς ένας τρισδιάστατος πίνακας (ύψος, πλάτος, κανάλια χρωμάτων).

Η είσοδος του Παραγωγικού Δικτύου είναι συνήθως ένας μονοδιάστατος τυχαίος (θόρυβος) πίνακας, ο οποίος θα ανακατασκευαστεί με τα αντίστοιχα στρώματα και στην έξοδο θα έχει φτάσει τις κατάλληλες διαστάσεις, όσες δηλαδή είναι και οι διαστάσεις των εικόνων στη συλλογή δεδομένων που χρησιμοποιήθηκαν πχ (64,64,3).

Η διαδικασία κατά την οποία ανακατασκευάζεται ένας πίνακας συνηθίζεται να γίνεται με δύο τεχνικές:

- Υπερδειγματοληπτικό Στρώμα (Upsampling) : είναι ένα στρώμα το οποίο χρησιμοποιείται για να αυξήσει-ανεβάσει διαστάσεις. Η διαδικασία που χρησιμοποιείται δεν απαιτεί εκπαίδευση, συνεπώς δεν διαθέτει βάρη ούτε μπορεί να χρησιμοποιηθεί για την εξαγωγή χαρακτηριστικών μιας εικόνας (feature extraction). Η αύξηση των διαστάσεων θα γίνει πάντοτε με καθορισμένο κριτήριο το οποίο μπορεί να είναι είτε ο πλησιέστερος γείτονας (nearest neighbor) ή η διγραμμική παρεμβολή (bilinear interpolation). Το στρώμα αυτό όπως αναφέρθηκε δεν μεταφέρει καμία πληροφορία και για αυτό το λόγο συνήθως συνδυάζεται με κάποιο συνελικτικό στρώμα (Convolutional Layers) το οποίο διαθέτει βάρη και επομένως πληροφορία για το μοντέλο.
- Ανεστραμμένο Συνελικτικό (Convolutional Transpose): Πραγματοποιεί την αντίστροφη διαδικασία του Συνελικτικού Στρώματος, πιο συγκεκριμένα η προώθηση των δεδομένων στα επόμενα στρώματα όπως και η διαδικασία οπισθοδρόμισης του Convolution Transpose είναι αντίστροφη στα συνελικτικά δίκτυα. Έχει σαν μεταβλητές το φίλτρο (filter) που αφορά τα κανάλια χρωμάτων μιας εικόνας, τον πυρήνα (kernel size) που αφορά τις διαστάσεις (ύψος, πλάτος) και αναλόγως την τιμή του βήματος (stride) κάνει την αύξηση των διαστάσεων. Στη πλειοψηφία των περιπτώσεων χρησιμοποιείται αυτό το στρώμα λόγω του ότι με μεγαλύτερη χωρική οικονομία προσφέρει πληροφορία στο μοντέλο.

```

1 def define_generator(latent_dim):
2     model = Sequential()
3     # foundation for 7x7 image
4     n_nodes = 128 * 7 * 7
5     model.add(Dense(n_nodes, input_dim=latent_dim))
6     model.add(LeakyReLU(alpha=0.2))
7     model.add(Reshape((7, 7, 128)))
8     # upsample to 14x14
9     model.add(Conv2DTranspose(128, (4,4), strides=(2,2),
10                           padding='same'))
11    model.add(LeakyReLU(alpha=0.2))
12    # upsample to 28x28
13    model.add(Conv2DTranspose(128, (4,4), strides=(2,2),
14                           padding='same'))
15    model.add(LeakyReLU(alpha=0.2))
# generate
model.add(Conv2D(1, (7,7), activation='tanh', padding='same'))
```

```
16     return model
```

Listing 2.1: Κώδικας απλού Παραγωγικού Δικτύου

Όπως παρατηρείται και στο παράδειγμα, η είσοδος είναι ένας μονοδιάστατος τένσορας (latent dim) και σταδιακά αυξάνονται οι διαστάσεις του.

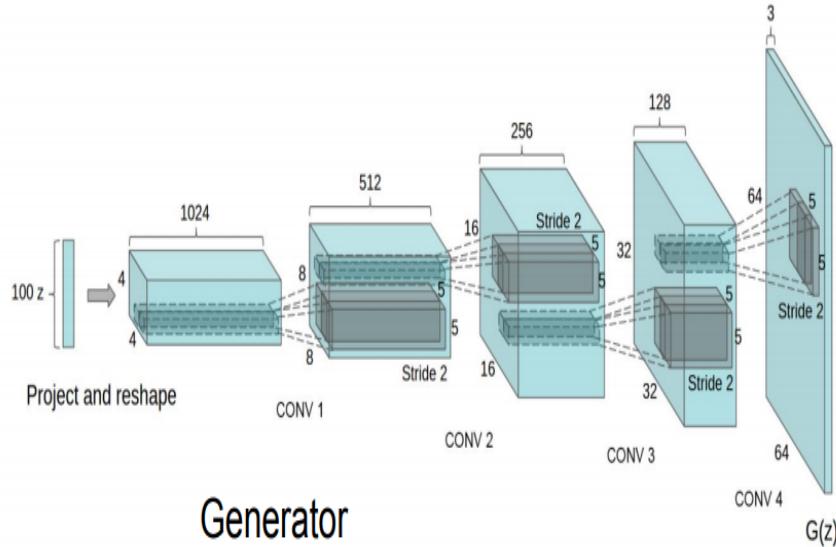


Figure 2.3: Δομή Παραγωγικού Δικτύου [2]

2.1.2 Διαχωριστικό Δίκτυο

Το Διαχωριστικό Δίκτυο έχει στόχο να ξεχωρίσει τις πλασματικές εικόνες του Παραγωγικού δικτύου από τις “αυθεντικές” εικόνες της αρχικής συλλογής δεδομένων. Η διαδικασία αυτή στον κλάδο της βαθιάς μάθησης πραγματοποιείται με τη ταξινόμηση (classification) δεδομένων σε δύο κλάσεις, στη πλασματική και στην αληθινή. Η δομή του Διαχωριστικού Δικτύου χαρακτηρίζεται από την αλληλουχία συνελικτικών στρωμάτων, όπως κάθε συνελικτικό νευρωνικό δίκτυο. Η είσοδος του Διαχωριστικού Δικτύου είναι μια εικόνα, δηλαδή ένας πολυδιάστατος πίνακας (ύψος, πλάτος, κανάλια χρωμάτων) και με τα συνελικτικά δίκτυα επιτυγχάνεται η σταδιακή μείωση των διαστάσεων της. Σαν συνάρτηση ενεργοποίησης χρησιμοποιείται η σιγμοειδής, λόγω των τελικών δύο κλάσεων.

```

1 def define_discriminator(in_shape=(28, 28, 1)):
2     model = Sequential()
3     # downsample
4     model.add(Conv2D(128, (3,3), strides=(2,2), padding='same',
5                      input_shape=in_shape))
6     model.add(LeakyReLU(alpha=0.2))
7     # downsample
8     model.add(Conv2D(128, (3,3), strides=(2,2), padding='same'))
9     model.add(LeakyReLU(alpha=0.2))
10    # classifier
11    model.add(Flatten())
12    model.add(Dropout(0.4))
13    model.add(Dense(1, activation='sigmoid'))
14    # compile model
15    opt = Adam(lr=0.0002, beta_1=0.5)
16    model.compile(loss='binary_crossentropy', optimizer=opt,
                     metrics=['accuracy'])
    return model

```

Listing 2.2: Κώδικας απλού Διαχωριστικού Δικτύου

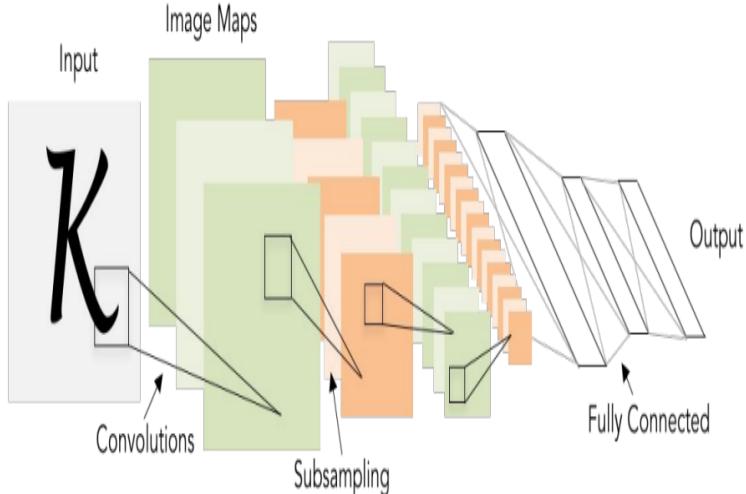


Figure 2.4: Δομή Διαχωριστικού Δικτύου[3]

2.1.3 Διαδικασία εκπαίδευσης

Ένα Παραγωγικό Αντιπαλικό Δίκτυο έχει ολοκληρωμένη δομή μόνο όταν ενωθεί η έξοδος του Παραγωγικού Δικτύου με τον είσοδο του Διαχωριστικού

Δικτύου. Ιδιαίτερο ενδιαφέρον έχει η διαδικασία εκπαίδευσης ενός Παραγωγικού Αντιπαλικού Δικτύου. Κατά τη διαδικασία αυτή, τα δύο τμήματα του ΠΑΔ ανταγωνίζονται μεταξύ τους, το Παραγωγικό με στόχο να “ξεγελάσει” το Διαχωριστικό, και το Διαχωριστικό με στόχο να μην “ξεγελαστεί”. Ο κύκλος εκπαίδευσης του ΠΑΔ χωρίζεται σε δύο μέρη: στο πρώτο μέρος ανήκει η αναβάθμιση των βαρών του Παραγωγικού Δικτύου και στο δεύτερο μέρος η αναβάθμιση των βαρών του Διαχωριστικού Δικτύου, ποτέ όμως ταυτόχρονα και τα δύο Δίκτυα - απαιτείται να πραγματοποιηθούν δύο ξεχωριστές οπισθοδρομήσεις (backpropagations). Ο Yann LeCun, κάτοχος βραβείου Turing για την συνεισφορά του στον τομέα της βαθιάς μάθησης έχει χαρακτηρίσει το Παραγωγικό Αντιπαλικό Δίκτυο ως ”την πιο έξυπνη ιδέα στη μηχανική μάθηση τα τελευταία είκοσι χρόνια” [4]. Στη συνέχεια περιγράφεται η διαδικασία που πρέπει να ακολουθήσει ένα σύστημα ΠΑΔ για να εκπαιδευτεί.

- Κατασκευή πλασματικών δεδομένων από το Παραγωγικό Δίκτυο και τοποθέτηση τους στην κλάση 0 ενώ τα ”αυθεντικά” δεδομένα τοποθετούνται στην κλάση 1.
- Κατασκευή ενός συνόλου δεδομένων, το οποίο αποτελείται από πλασματικά (κλάση 0) και ”αυθεντικά” (κλάση 1) δεδομένα. Προτιμάται η επιλογή ενός συνόλου που να περιέχει και τις δύο κλάσεις από ότι δύο ξεχωριστών συνόλων [51].
- Χρησιμοποιείται αυτό το σύνολο δεδομένων για να εκπαιδευτεί το Διαχωριστικό Δίκτυο. Στην περίπτωση δύο ξεχωριστών συνόλων πραγματοποιείται δύο φορές η εκπαίδευση.
- Στη συνέχεια, όντας εκπαιδευμένο το Διαχωριστικό Δίκτυο, χρησιμοποιείται όλο το ΠΑΔ μοντέλο που έχει και το Παραγωγικό και το Διαχωριστικό Δίκτυο και απαιτείται να ”παγώσει” το κομμάτι του Διαχωριστικού Δικτύου έτσι ώστε η οπισθοδρόμηση να μην επηρεάσει τα βάρη του.
- Ξεκινάει η εκπαίδευση του Παραγωγικού Δικτύου, κατά την οποία κατασκευάζονται πλασματικά δεδομένα αλλά τοποθετούνται στην κλάση 1, δηλαδή στα ”αυθεντικά” δεδομένα. Το σκεπτικό πίσω από αυτή την πρακτική είναι ότι, αν το Διαχωριστικό Δίκτυο προβλέψει ότι είναι πλασματικά τα δεδομένα ενώ έχουν τοποθετηθεί στη κλάση των ”αυθεντικών”, θα οδηγηθεί το μοντέλο σε μεγάλο σφάλμα που με τη σειρά του θα οδηγήσει σε μεγάλες αλλαγές και αναβάθμισεις στα βάρη του Παραγωγικού Δικτύου, με στόχο να βελτιωθεί και να παράξει πιο ρεαλιστικά δεδομένα.

Σημαντική παρατήρηση αποτελεί το γεγονός ότι σε κάθε νέο κύκλο εκπαίδευσης του ΠΑΔ, το Διαχωριστικό Δίκτυο πρέπει να "ξεπαγώνει" και να "παγώνει" την κατάλληλη χρονική στιγμή.

Κατά τη διαδικασία εκπαίδευσης ενός ΠΑΔ, αρκετά φαινόμενα, που μπορούν να προβληματίσουν, μπορούν να κάνουν την εμφάνισή τους. Πρόκειται για δύο ταυτόχρονες εκπαιδεύσεις, οι οποίες ανταγωνίζονται μεταξύ τους, δηλαδή η βελτίωση του ενός τμήματος θα σημαίνει μεγαλύτερο σφάλμα στο άλλο τμήμα.

Δύο βασικά προβλήματα συναντάει κάποιος στην εκπαίδευση του ΠΑΔ:

- **Βλάβη Μοντέλου (Model Collapse [5]):** Μια επιτυχημένη εκπαίδευση οδηγεί σε ένα Παραγωγικό Δίκτυο το οποίο έχει τη δυνατότητα να παράξει πολλά διαφορετικά πλασματικά δεδομένα. Παρόλα αυτά, στόχος του Παραγωγικού Δικτύου είναι να παράξει δεδομένα τα οποία να είναι όσο το δυνατό πιο ρεαλιστικά για το Διαχωριστικό Δίκτυο και μόνο. Πολλές φορές ένα ή ελάχιστα τέτοια δεδομένα αρκούν στο Παραγωγικό Δίκτυο με αποτέλεσμα να παράγει μόνο αυτά. Η κατάσταση αυτή φαίνεται όταν θα χρησιμοποιηθεί το Παραγωγικό Δίκτυο, το οποίο θα μπορέσει να παράξει μόνο ένα περιορισμένο πλήθος διαφορετικών εικόνων. Φαίνεται επίσης στο διάγραμμα σφάλματος του Παραγωγικού Δικτύου καθώς παρουσιάζει ταλαντώσεις. Οι ταλαντώσεις οφείλονται στις καταστάσεις που βρίσκεται το Παραγωγικό Δίκτυο, αρχικά χρησιμοποιεί ένα πλήθος εικόνων το οποίο δεν μπορεί να αναγνωρίσει το Διαχωριστικό Δίκτυο, ύστερα το Διαχωριστικό Δίκτυο καταφέρνει να τις ταξινομήσει ορθά και το Παραγωγικό Δίκτυο αναγκάζεται να βρει ένα νέο γκρουπ εικόνων, πιο περιορισμένο το οποίο να μπορέσει πάλι να "ξεγελάσει" το Διαχωριστικό Δίκτυο. Συνήθως οφείλεται στην δομή του μοντέλου.
- **Αποτυχία Σύγκλισης (Non Convergence [5]):** Καθώς το Παραγωγικό Δίκτυο βελτιώνεται, η ακρίβεια του Διαχωριστικού Δικτύου πρέπει να πλησιάζει το 50%, δηλαδή να είναι τυχαία η επιλογή του καθώς δεν θα μπορέσει να διακρίνει την διαφορά "ψεύτικων"- "αυθεντικών" δεδομένων. Όταν το μοντέλο αδυνατεί να φτάσει σε αυτή την κατάσταση ισορροπίας - συνήθως το βλέπουμε όταν η ακρίβεια του Διαχωριστικού Δικτύου αυξάνεται όπως και το σφάλμα του Παραγωγικού Δικτύου - εμφανίζεται το φαινόμενο της αποτυχίας σύγκλισης. Το Παραγωγικό Δίκτυο παράγει δεδομένα όχι καλής ποιότητας, τα οποία το Διαχωριστικό Δίκτυο μπορεί να τα καταλάβει και να ταξινομήσει εύκολα.

Τα παραπάνω προβλήματα μπορεί να οφείλονται σε διάφορους λόγους, από την κατανομή και τη φύση των δεδομένων ή την λανθασμένη αρχιτεκτονική του μοντέλου. Κάποιες τεχνικές που μπορούν να βοηθήσουν σε μια πιο σταθερή εκπαίδευση ενός ΠΑΔ είναι οι εξής :

- Εξομάλυνση Κλάσεων (Label Smoothing): Σύμφωνα με τα παραπάνω το 1 αντιστοιχεί στα "αυθεντικά" και το 0 στο πλασματικά δεδομένα. Η εξομάλυνση κλάσεων προτείνει η κλάση των "αυθεντικών" δεδομένων να μην είναι αυστηρά το 1 αλλά να κυμαίνεται σε μια τιμή γύρω από το 1 και η κλάση των πλασματικών σε μια τιμή γύρω από το 0. Ο λόγος που οδηγεί σε μια πιο σταθερή εκπαίδευση είναι γιατί πλέον ενισχύεται η γενίκευση (generalization), καιώς μειώνεται η "αυτοπεποίθηση" του μοντέλου μας. [58]
- Χρησιμοποιούνται Ανεστραμμένα Συνελικτικά στρώματα με βήμα αντί για Υπερδειγματοληπτικά (Upsampling) στρώματα
- Θόρυβο στο Διαχωριστικό Δίκτυο [6]: Επιτυγχάνεται με αντιστροφή κλάσεων, δηλαδή μερικά "αυθεντικά" δεδομένα τοποθετούνται στην κλάση 0 και το αντίστροφο.
- Υπερβολική Εφαπτομένη (tanh) σαν συνάρτηση ενεργοποίησης στο Παραγωγικό Δίκτυο: Μια φραγμένη συνάρτηση ενεργοποίησης οδηγεί στο μοντέλο να μάθει πιο γρήγορα στο πως να καλύψει τα χρώματα σε μια εικόνα. Για αυτό το λόγο πραγματοποιείται μια κλιμάκωση στα δεδομένα στο εύρος [-1,1] πριν ξεκινήσει η διαδικασία εκπαίδευσης [56].
- Adam Optimizer: Με τις υπερπαραμέτρους learning rate = 0.0002 και beta1 momentum = 0.5.
- Στρώμα "διαφροής" διορθωμένης γραμμικής μονάδας (Leaky Relu) ύστερα από κάθε συνελικτικό ή ανεστραμμένο συνελικτικό στρώμα.
- Αρχικοποίηση Βαρών (Weight Initialization): Πιο συγκεκριμένα, χρησιμοποιείται κανονική κατανομή με μέση τιμή 0 και διακύμανση 0.02 [56].

Τα κριτήρια που πρέπει να αξιολογηθούν για να αποφανθεί μια αποτελεσματική και σταθερή εκπαίδευση είναι οι τιμές του σφάλματος και της ακρίβειας του Παραγωγικού και Διαχωριστικού Δικτύου, όπως επίσης και το τελικό οπτικό αποτέλεσμα, το οποίο πρέπει να είναι αρκετά πειστικό.

2.1.4 Διάφορες Χρήσεις Παραγωγικών Αντιπαλικών Δικτύων

DCGAN [56]: Deep Convolutional GAN (Βαθύ Συνελικτικό ΠΑΔ), το οποίο έχει αντικείμενο τις εικόνες και έχει την δομή που περιγράφεται παραπάνω. Στα πλαίσια αυτής της διπλωματικής το ΠΑΔ που χρησιμοποιείται ανήκει σε αυτή την κατηγορία. Αρκετά ενδιαφέρον αποτελεί το γεγονός ότι το DCGAN έχει

αρκετές χρήσεις και αποτελεί εξαιρετικό εργαλείο έπειτα από μια πετυχημένη εκπαίδευση. Αρχικά, το Παραγωγικό Δίκτυο μπορεί να χρησιμοποιηθεί για να παράξει νέα δεδομένα και να ενισχύσει μια μικρή συλλογή δεδομένων με νέες εικόνες με ρεαλιστικά χαρακτηριστικά. Παράλληλα το Διαχωριστικό Δίκτυο είναι ένας εκπαιδευμένος ταξινομητής, ο οποίος σε κάθε στάδιο μείωσης των διαστάσεων των εικόνων φέρει πληροφορίες για τα χαρακτηριστικά τους. Με την τεχνική της Μεταφοράς Γνώσης (Transfer Learning) μπορούν να αξιοποιηθούν αυτές οι πληροφορίες σε παρόμοιες συλλογές δεδομένων.

CycleGAN [68]: Χρησιμοποιούνται για να λύσουν το πρόβλημα Μεταφοράς Εικόνας σε Εικόνα (Image-to-Image Translation). Έχοντας δύο διαφορετικές συλλογές δεδομένων, κάθε μια με συγκεκριμένο πεδίο, επιτυγχάνεται η μεταφορά του στυλ κάθε πεδίου στο πεδίο της αντίστοιχης συλλογής δεδομένων. Για παράδειγμα αν το πρώτο πεδίο αφορά χειμωνιάτικα τοπία ενώ το δεύτερο πεδίο καλοκαιρινά τοπία, μπορούν οι φωτογραφίες των τοπίων του χειμώνα να εμφανιστούν για καλοκαίρι. Η δομή και το σκεπτικό της εκπαίδευσης ενός CycleGAN είναι η εξής:

- Δύο ζευγάρια Παραγωγικών-Διαχωριστικών Δικτύων (G1-D1,G2-D2)
- Εικόνες από το πεδίο-1 μπαίνουν στον G1 και το αποτέλεσμα του προσπαθεί να το ξεχωρίσει ο D1 σχέση με τις εικόνες από το πεδίο-2
- Παράλληλα η έξοδος του G1 μπαίνει στον G2 και την έξοδο του G2 προσπαθεί να την ξεχωρίσει ο D2 σε σχέση με το πεδίο-2



Figure 2.5: Έξοδος CycleGAN [7]

InfoGAN [26]: Η συγκεκριμένη παραλλαγή, δίνει τη δυνατότητα να υπάρχει έλεγχος στα χαρακτηριστικά της εικόνας που θα παράξει το Παραγωγικό Δίκτυο. Αυτό συμβαίνει αν επεκταθεί η δομή του DCGAN και πλέον το Παραγωγικό Δίκτυο εκτός από τυχαίο αρχικό θόρυβο αποκτήσει σαν είσοδο και έναν "κωδικό". Αυτός ο κωδικός μπορεί είτε να αντιστοιχεί σε μια χλάση και να έχει τη μορφή του one hot vector ή να είναι συνεχής και να παράγεται από μια κατανομή. Τότε το μοντέλο συσχετίζει κάθε κωδικό με τα διαφορετικά χαρακτηριστικά μιας εικόνας, π.χ. αν χρησιμοποιούνται ως συλλογή δεδομένων πρόσωπα, τότε μια συγκεκριμένη κατανομή (δηλαδή κατανομή με συγκεκριμένη μέση τιμή και απόκλιση) μπορεί να οδηγεί σε διαφορετικό χρώμα μαλλιών.

2.2 Μεταφορά Γνώσης

Στις παραδοσιακές προσεγγίσεις της Μηχανικής Μάθησης, κάθε μοντέλο κατασκευάζεται έτσι ώστε να αντιμετωπίζει ένα συγκεκριμένο σετ δεδομένων. Αυτό έχει ως αποτέλεσμα να μην εκμεταλλεύεται η πληροφορία όλων των προηγούμενων μοντέλων που έχουν εκπαιδευτεί αλλά και να στηρίζεται η εκπαίδευση των νέων μοντέλων σε πολλούς πόρους και στην εύρεση επαρκούς πλήθους νέων δεδομένων.

Η Μεταφορά Γνώσης είναι η τεχνική κατά την οποία εκπαιδεύονται μοντέλα με ένα σετ δεδομένων και ύστερα αυτά τα εκπαιδευμένα μοντέλα αξιοποιούνται σε

διαφορετικά αλλά παρεμφερή δεδομένα. Όταν ένα μοντέλο εκπαιδεύεται πάνω σε ένα σετ δεδομένων, τότε προσπαθεί να εξάγει πληροφορίες για την κατανομή του. Όλη αυτή η πληροφορία είναι χρήσιμη και δεν πρέπει να αγνοηθεί, καθώς μπορεί να αξιοποιηθεί άμεσα χωρίς εκ νέου εκπαίδευση σε ένα νέο σετ δεδομένων. Σε περιπτώσεις που διατίθενται λίγοι πόροι ή ελάχιστα δεδομένα η τεχνική της Μεταφοράς Γνώσης μπορεί να δώσει εξαιρετικά αποτελέσματα. Η έννοια αυτή εισήχθη πρώτη φορά στο συνέδριο NIPS το 1995 στην ημερίδα "Learning to Learn: Knowledge Consolidation and Transfer in Inductive Systems" [8] και άρχισε να εξελίσσεται ιδιαίτερα την τελευταία δεκαετία.

2.2.1 Κατηγορίες Μεταφοράς Γνώσης στο πεδίο της Μηχανικής Μάθησης

Για να αναλυθούν οι διάφορες τεχνικές ωστε πρώτα να εισαχθούν δύο έννοιες: η πηγή (Source) και ο στόχος (Target). Η πηγή, είναι διαδικασία της πρωταρχικής εκπαίδευσης του μοντέλου. Η πηγή διαθέτει το πεδίο (domain) της, το οποίο είναι το σετ δεδομένων της και έχει και το ζητούμενο (task) το οποίο είναι το τι ακριβώς ζητείται να επιτευχθεί με αυτά τα δεδομένα (πχ ταξινόμηση, αναγνώριση, εντοπισμός). Αντίστοιχα υπάρχει ο στόχος, ο οποίος αφορά την δεύτερη εκπαίδευση, την εκπαίδευση όπου ωστε αξιοποιηθούν οι πληροφορίες που δίνει η πρώτη εκπαίδευση, και διαθέτει το δικό του πεδίο και ζητούμενο, π.χ. για την οδήγηση μηχανής (στόχος), μπορεί να χρησιμοποιηθεί η γνώση οδήγησης ποδηλάτου (πηγή).

Γενικές Κατηγορίες Μεταφοράς Γνώσης στο πεδίο της Μηχανικής Μάθησης:

- Μη επιβλεπόμενη (Unsupervised): Σε αυτή την περίπτωση δεν υπάρχουν ταξινομημένα δεδομένα (labeled data) ούτε στη πηγή ούτε στο στόχο. Ανάλογα με το ζητούμενο της πηγής και του στόχου ωστε γίνει προσπάθεια να μεταφερθεί η γνώση από την κατανομή των δεδομένων του πεδίου της πηγής στο ζητούμενο του στόχου.
- Μεταγωγική (Transductive): Σε αυτή την περίπτωση υπάρχουν ταξινομημένα δεδομένα μόνο στο πεδίο της πηγής
- Επαγωγική (Inductive): Σε αυτή την περίπτωση υπάρχουν ταξινομημένα δεδομένα σίγουρα στο πεδίο του στόχου (ίσως και στο πεδίο της πηγή)

2.2.2 Τεχνικές και Μέθοδοι Μεταφοράς Γνώσης στο πεδίο της Βαθιάς Μάθησης

Τα μοντέλα στο πεδίο της Βαθιάς Μάθησης βασίζονται στην ιδιότητα ότι καταφέρνουν να δημιουργήσουν μια αναπαράσταση για τα δεδομένα εισόδου τους. Στόχος της Μεταφοράς Γνώσης στο πεδίο της Βαθιάς Μάθησης είναι να μπορέσει να μεταφερθεί αυτή η αναπαράσταση και να εφαρμοστεί στο ζητούμενο του στόχου. Δεδομένου ότι υπάρχει ένα μοντέλο που έχει εκπαίδευτεί με ένα πεδίο πηγής, δηλαδή ένα σετ δεδομένων, έχουν αναπτυχθεί ορισμένες τεχνικές για την μεταφορά της αναπαράστασης, όπως αυτές παρουσιάζονται στη συνέχεια.

- **Αντικατάσταση Τελικού Στρώματος (Final Layer Replacement):** Κάθε μοντέλο Βαθιάς μάθησης αποτελείται από πολλαπλά στρώματα, το καθένα από αυτά χρησιμεύει για να εξάγει διάφορα χαρακτηριστικά των δεδομένων, συνήθως όσο πιο προχωρημένο είναι ένα στρώμα (δηλαδή βρίσκεται στο τέλος της δομής του μοντέλου) τόσο πιο σύνθετη θα είναι και η πληροφορία που θα έχει συλλέξει. Με την τεχνική αυτή αντικαθίσταται το τελευταίο στρώμα το οποίο είναι ένα πλήρως συνδεδεμένο (fully connected) και είναι υπεύθυνο για την έξοδο και ικανοποίηση του ζητούμενου της πηγής. Στη θέση του, τοποθετείται ένα στρώμα που θα εξυπηρετεί το ζητούμενο του στόχου.
- **Αντικατάσταση Στρωμάτων (Layers Replacement) :** Με το ίδιο σκεπτικό, όσο λιγότερα χαρακτηριστικά επιλεχθούν να αξιοποιηθούν από την πηγή τόσα περισσότερα στρώματα θα αντικατασταθούν, από το τέλος προς την αρχή της δομής του μοντέλου

Έχοντας πλέον την τελική αρχιτεκτονική του μοντέλου έτοιμη με τις αντικαταστάσεις στρωμάτων υπάρχει η δυνατότητα να εφαρμοστούν δύο τεχνικές κατά την διάρκεια της εκπαίδευσης

- **”Πάγωμα” Στρωμάτων (“Freezing” Layers) :** Γίνεται επιλογή των στρωμάτων, τα οποία δεν είναι επιθυμητό να αναβαθμίσουν τα βάρη τους με την οπισθοδόμηση που θα προκαλέσει η εφαρμογή του σετ δεδομένων του στόχου, και ”παγώνονται”. Με αυτόν τον τρόπο διατηρείται όλη η πληροφορία που έχουν συλλέξει από το σετ της πηγής.
- **Ρύθμιση Στρωμάτων-Επανεκπαίδευση (Fine Tuning):** Από την αντίθετη πλευρά, τα στρώματα που έχουν την δυνατότητα να αναβαθμίσουν τα βάρη τους, συλλέγουν πληροφορίες από το σετ του στόχου και διαμορφώνουν τα βάρη τους ανάλογα.

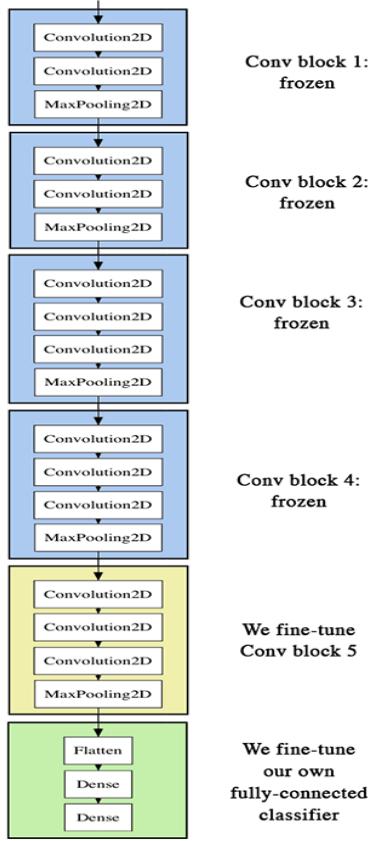


Figure 2.6: Μοντέλο Μεταφοράς Γνώσης [9]

Όπως συνεπάγεται από τα παραπάνω ένα προεκπαίδευμένο μοντέλο που δίνει τη δυνατότητα σε όλα τα στρώματα του να εκπαιδευτούν ξανά με το πεδίο του στόχου, είναι η ιδανική περίπτωση. Όμως τις περισσότερες φορές δεν μπορεί να εφαρμοστεί αυτή η τεχνική λόγω των απαιτήσεων που θα έχει ένα τέτοιο μοντέλο σε υπολογιστικούς πόρους. Το trade-off μεταξύ επίδοσης και πόρων πρέπει να αξιολογηθεί και να χειρίστει ξεχωριστά σε κάθε κατάσταση.

Μια μέθοδος που χρησιμοποιείται για να εξοικονομήσει υπολογιστικό χρόνο και πόρους λειτουργεί με αποκοπή των "παγωμένων" στρωμάτων από το υπόλοιπο μοντέλο. Σαν είσοδο του αποκομμένου συστήματος τοποθετείται το σετ του στόχου και γίνεται συλλογή της εξόδου. Πλέον η έξοδος αυτή είναι το νέο σετ που θα χρησιμοποιηθεί σαν είσοδος στο υπόλοιπο εκπαιδεύσιμο μοντέλο. Με τον τρόπο αυτό το αποκομμένο σύστημα εκτελείται μόνο μια φορά, ενώ στην αντίθετη περίπτωση τα δεδομένα θα περνούσαν πολλές φορές ανάλογα με τις εποχές που θα

ορίζαμε.

2.2.3 Μεταφορά Γνώσης στο πεδίο της Υπολογιστικής Όρασης

Από το 2010 έως το 2017 υπήρχε ένας ετήσιος διαγωνισμός, ο ImageNet Large Scale Visual Recognition Challenge ή ILSVRC, ο οποίος είχε ως ζητούμενο την ταξινόμηση εικόνων (image classification) και τον εντοπισμό αντικειμένων σε εικόνες (object detection). Ως πεδίο πηγής χρησιμοποιούταν το ImageNet Dataset[10] ή ένα μέρος του. Το ImageNet Dataset αποτελείται από περισσότερες από 14 εκατομμύρια εικόνες και πάνω από 20 χιλιάδες κατηγορίες-κλάσεις (labels). Ο διαγωνισμός αυτός είχε ως αποτέλεσμα να αναπτυχθεί σημαντικά ο τομέας της υπολογιστικής Όρασης και να βελτιωθούν οι state of the art τεχνικές.

Classification Results (CLS)

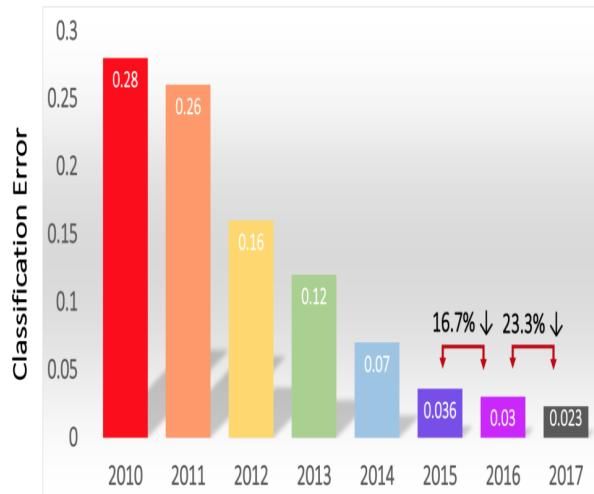


Figure 2.7: Error rate [11]

Η μεγάλη πτώση του σφάλματος ταξινόμησης συνέβη το 2012 όπου έχουμε σχεδόν 10% μείωση του σφάλματος. Υπεύθυνο μοντέλο για αυτή την επίδοση είναι το AlexNet, το οποίο βασίζεται σε συνελικτικά στρώματα αλλά και στη χρήση της GPU για την εκπαίδευση του.

Πλέον, το ImageNet Dataset έχει χρησιμοποιηθεί για να εκπαιδεύσει πολλά προεκπαίδευμένα μοντέλα τα οποία παρέχονται στα διάφορα frameworks βαθιάς

μάθησης με την αντίστοιχη αρχιτεκτονική τους και διευκολύνουν πολύ την διαδικασία της Μεταφοράς Γνώσης. Μερικά από αυτά τα μοντέλα είναι τα εξής:

- VGG
- ResNet
- DenseNet
- EfficientNet
- Inception

Εκτός από αυτά τα προεκπαιδευμένα μοντέλα, πλέον υπάρχουν μοντέλα πιο εξειδικευμένα σε διαφορετικά ζητούμενα πέρα από τον τομέα της υπολογιστικής όρασης πχ επεξεργασία φυσικής γλώσσας, και παρέχονται δωρεάν στο διαδίκτυο μαζί με τα αντίστοιχα βάρη τους.

2.3 Επαύξηση Δεδομένων

Στο πεδίο της μηχανικής μάθησης και της βαθιάς μάθησης, τα δεδομένα ίσως είναι ο πιο σημαντικός παράγοντας για την αποτελεσματικότητα του συστήματος. Ένα μοντέλο κατά τη διαδικασία εκπαίδευσης χρειάζεται δεδομένα έτσι ώστε να αντιληφθεί το μοτίβο, τις αλληλοσυσχετίσεις και την κατανομή με στόχο να οδηγηθεί σε καλύτερη ακρίβεια και μειωμένο σφάλμα. Ένα τρανταχτό παράδειγμα της "δύναμης" των δεδομένων είναι τα μοντέλα που πραγματεύονται κείμενα και έχουν στόχο την επεξεργασία φυσικής γλώσσας, ένα ερευνητικό πεδίο όπου υπάρχει αισθητή βελτίωση τα τελευταία χρόνια (π.χ. μετάφραση κειμένων).

Πάρα πολλές φορές χρειάζεται να αντιμετωπιστεί κάποια συλλογή με λίγα δεδομένα ή με δεδομένα που παρουσιάζουν άνιση κατανομή. Η συλλογή δεδομένων και η συνεχής ενίσχυση των υπαρχόντων σετ προϋποθέτει υπολογιστικούς πόρους (για την εξόρυξη τους) ή οικονομικό κόστος (για την αγορά τους). Μια μέθοδος που μπορεί να δώσει μια μερική λύση σε αυτό το πρόβλημα είναι η Επαύξηση Δεδομένων. Η μέθοδος αυτή βρίσκει τεράστια εφαρμογή στον τομέα της υπολογιστικής όρασης.

2.3.1 Επαύξηση Δεδομένων στην Υπολογιστική Όραση

Ως δεδομένο στην υπολογιστική όραση θεωρείται κάθε είδους εικόνα. Η Επαύξηση Δεδομένων, είναι η διαδικασία κατά την οποία ενισχύεται το σετ εικόνων

που διατίθεται ως δεδομένα, χωρίς να γίνει συλλογή νέων εικόνων. Αυτό επιτυγχάνεται με τις διάφορες τροποποιήσεις στις ήδη υπάρχουσες εικόνες με σκοπό την επαύξηση του σετ στο οποίο θα χρησιμοποιήσει ο αλγόριθμος μηχανικής μάθησης. Πιο συγκεκριμένα οι πιο δημοφιλείς επιλογές για επαύξηση δεδομένων είναι:

- Περιστροφή: γωνία περιστροφής
- Οριζόντια/Κάθετη Αναστροφή: True/False
- Οριζόντια/Κάθετη Μετατόπιση: Pixels

Τπάρχουν και επιπλέον επιλογές όπως `brightness_range`, `rescale`, `crop`, `zoom_range` κ.α.

Κάθε επιλογή που γίνεται για την τροποποίηση των εικόνων, πρέπει να συμβαδίζει με τη φύση των εικόνων που περιμένει το μοντέλο. Για παράδειγμα, θα εμπλουτίσει το σετ δεδομένων το να εφαρμοστεί οριζόντια αναστροφή σε ένα τοπίο αλλά δεν έχει νόημα να εφαρμοστεί κάθετη αναστροφή σε ένα σκύλο.



Figure 2.8: Περιστροφή και Οριζόντια Αναστροφή

2.3.2 Επαύξηση Δεδομένων και Πλεονεκτήματα

Τα βασικά ζητήματα τα οποία η Επαύξηση Δεδομένων μπορεί να επηρεάσει θετικά είναι τα εξής:

Τπερπροσαρμογή (Overfitting): Κατά το φαινόμενο αυτό, το μοντέλο μαθαίνει και προσαρμόζεται αποκλειστικά σε μια κατανομή δεδομένων, με αποτέλεσμα να αποδίδει πολύ καλά σε αυτήν, αλλά να εμφανίζει κακή απόδοση σε οποιαδήποτε άλλη κατανομή έχει σαν είσοδο. Αυτό το φαινόμενο μπορεί να προκαλέσει πρόβλημα όταν το σετ δεδομένων είναι μη ισορροπημένο, και το μοντέλο μαθαίνει πολύ καλά την κλάση με το μεγαλύτερο πλήθος εικόνων. Επίσης παρατηρείται σε περιπτώσεις που χρησιμοποιείται ένας μεγάλος αριθμός εποχών αλλά μια μικρή συλλογή δεδομένων για την εκπαίδευση. Με αυτόν τον τρόπο, το μοντέλο μαθαίνει τα δεδομένα

εκπαίδευσης αλλά αδυνατεί να αποδώσει στα δεδομένα ελέγχου ή γενικότερα σε ξένα δεδομένα. Με την Επαύξηση Δεδομένων μπορεί να διορθωθεί αυτό το πρόβλημα είτε με αύξηση των εικόνων της κλάσης με το μικρότερο πλήθος, διαδικασία που ονομάζεται υπερδειγματοληψία (oversampling), είτε με αύξηση του συνολικού σετ δεδομένων.

Ακρίβεια (Accuracy): Με την Επαύξηση Δεδομένων παρατηρείται αισθητή βελτίωση της ακρίβειας ενός μοντέλου μηχανικής μάθησης το οποίο έρχεται έμμεσα από τη δυνατότητα του να γενικεύει (generalize) καλύτερα. [53]

Αποθηκευτικός Χώρος: Υπάρχουν δύο τεχνικές υλοποίησης της Επαύξησης Δεδομένων, η ασύγχρονη (offline) και η σύγχρονη (online). Με την ασύγχρονη τεχνική πραγματοποιούνται οι τροποποιήσεις οι οποίες αποθηκεύονται και έτσι δίνεται η δυνατότητα να χρησιμοποιηθούν σε διαφορετικό χρόνο. Με την σύγχρονη, αφού πραγματοποιηθούν οι τροποποιήσεις, αυτές σχηματίζουν σύνολα τα οποία χρησιμοποιούνται άμεσα για να εκπαιδευτεί το μοντέλο, χωρίς να χρειάζεται κάποιος αποθηκευτικός χώρος για την μετέπειτα φύλαξη τους.

2.4 Βελτιστοποίηση Συμήνους Σωματιδίων

2.4.1 Εισαγωγή στον Αλγόριθμο

Η Βελτιστοποίηση Συμήνους Σωματιδίων (ΒΣΣ) [39] είναι ένας αλγόριθμος βελτιστοποίησης ο οποίος προτάθηκε το 1995 από τον James Kennedy και τον Russell C. Eberhart και εμπνεύστηκε από τη συμπεριφορά ζώων να σχηματίζουν ομάδα ή σμήνος και να μοιράζονται πληροφορίες μέσα σε αυτή, με σκοπό την καλύτερη επιβίωση τους. Στόχος της ΒΣΣ είναι να βρει τις κατάλληλες τιμές των διοθέτων μεταβλητών - στην συγκεκριμένη περίπτωση είναι υπερπαράμετροι - για τις οποίες επιτυγχάνεται η βέλτιστη απόδοση, δηλαδή μέγιστη ή ελάχιστη -όταν η μετρική είναι η απώλεια- τιμή μιας συνάρτησης εφαρμογής (fitness) $f(x)$.

Ο αλγόριθμος ΒΣΣ μπορεί να χαρακτηριστεί σαν μια αναζήτηση σε έναν n -διάστατο χώρο, όπου ν είναι ο αριθμός των υπερπαραμέτρων προς βελτίωση. Για να αναλυθεί το μαθηματικό υπόβαθρο της μεθόδου, παρατίθενται κάποια χαρακτηριστικά της αναζήτησης. Αρχικά η αναζήτηση γίνεται από έναν αριθμό σωματιδίων που ορίζεται εξαρχής. Για κάθε σωματίδιο πρέπει να σημειώνονται:

- οι υπερπαράμετροι που οδήγησαν στην καλύτερη ατομική επίδοση P_{best}
- Οι υπερπαράμετροι που οδήγησαν στην καλύτερη ομαδική επίδοση G_{best} (πριν από κάθε κίνηση τα σωματίδια επικοινωνούν έτσι ώστε να ενημερώσουν για κάποια πιθανή αλλαγή της G_{best}).
- η ταχύτητα (velocity) του, όταν σημειώθηκαν αυτές οι υπερπαράμετροι

Για να κατανοηθεί μαθηματικά ο αλγόριθμος ΒΣΣ, θα πρέπει να ορίστούν κάποια μεγέθη. Αρχικά ορίζεται ένα διάνυσμα από τις υπερπαραμέτρους a_i, b_i, c_i, \dots που τίθενται προς βελτιστοποίηση

$$\vec{X}^t = [a_i, b_i, c_i, \dots]$$

Κάθε σωματίδιο έχει ένα δικό του "προσωπικό" διάνυσμα με τυχαία αρχικοποίηση και ξεχωριστή εξέλιξη στις τιμές των υπερπαραμέτρων, το οποίο δείχνει τη θέση του στον πολυδιάστατο χώρο αναζήτησης. Κάθε σωματίδιο έχει επίσης μια δικιά του ταχύτητα V^t η οποία ορίζεται ως εξής :

$$V^{t+1} = w * V^t + c_1 r_1 * (Pbest - X^t) + c_2 r_2 * (Gbest - X^t) \quad (2.1)$$

Η νέα θέση κάθε σωματιδίου ανανεώνεται με τον εξής τρόπο :

$$X^{t+1} = X^t + V^{t+1} \quad (2.2)$$

Ο αλγόριθμος τερματίζει όταν ικανοποιηθεί το κριτήριο τερματισμού, το οποίο ονομάζεται T_{max} και δηλώνει τον μέγιστο αριθμό αναζητήσεων που θα πραγματοποιήσουν. Υπάρχουν και άλλα λιγότερα διαδεδομένα κριτήρια όπως μείωση διαφοράς στη βελτιστοποίηση ή μειωμένη κίνηση στο χώρο.

2.4.2 Παράμετροι Αλγορίθμου

Όπως παρατηρείται από την (2.1), ο αλγόριθμος έχει τις εξής παραμέτρους :

- w : συντελεστής αδράνειας (inertia weight) $\in [w_{min}, w_{max}]$
- r_1, r_2 : τυχαίοι αριθμοί $\in [0, 1]$
- c_1, c_2 : σταθερές επιτάχυνσης (acceleration coefficients)

Ο όρος $w * V^t$ ονομάζεται αδράνεια (inertia) και διαισθητικά είναι ο όρος που τείνει να διατηρήσει την προηγούμενη ταχύτητα κάθε σωματιδίου

Ο όρος $c_1 r_1 * (Pbest - X^t)$ ονομάζεται cognitive ή individual component στην παγκόσμια βιβλιογραφία και αυτός ο όρος τείνει να κινήσει το κάθε σωματίδιο κοντά στο Pbest του. Αυτό οφείλεται στο γεγονός ότι αν το σωματίδιο βρεθεί μακριά από το Pbest τότε η μετακίνηση του για να επιστρέψει θα πρέπει να είναι μεγάλη, το οποίο θα ικανοποιηθεί από τον όρο $Pbest - X^t$ που έχει μεγαλώσει.

Ο όρος $c_2 r_2 * (Gbest - X^t)$ ονομάζεται social component και κρατάει κάθε σωματίδιο σε επαφή με το Gbest, καθώς αν το σωματίδιο απομακρυνθεί από το Gbest ο όρος $(Gbest - X^t)$ μεγαλώνει έτσι ώστε η επόμενη του μετακίνηση να είναι αρκετά μεγάλη για να επιστρέψει.

Algorithm 1: Βέλτιστοποίηση Συγκριτιδίων

```

for each particle do
    Initialize X randomly
    Initialize V randomly
    Evaluate the fitness  $f(X)$ 
    Initialize Pbest with a copy of X
end
Initialize Gbest with a copy of the best X
 $t = 0$ 
while  $t < T_{max}$  do
    for each particle do
        Update V according to (2.1)
        Update X according to (2.2)
        Evaluate the fitness  $f(X)$ 
         $Pbest \leftarrow X$  if  $f(Pbest) < f(X)$ 
         $Gbest \leftarrow X$  if  $f(Gbest) < f(X)$ 
    end
end

```

Οι σταθερές επιτάχυνσης c_1, c_2 , υποδηλώνουν πόσο ομαδική ή ατομική δουλειά θα γίνει. Χαρακτηριστικά αν γίνει η υπόθεση ότι $c_1 = 0$ και $c_2 > 0$, τότε όλα τα σωματίδια ενοποιούνται και δεν εξερευνούν τοπικές βέλτιστες επιδόσεις. Αντίστοιχα αν $c_2 = 0$ και $c_1 > 0$ τότε χάνεται η έννοια της ομαδικότητας και δημιουργείται μια ατομική αναζήτηση για κάθε σωματίδιο στον δικό του τοπικό χώρο. Είναι σημαντικό να υπάρχει μια ισορροπία στις δύο αυτές παραμέτρους με μια καλή αναλογία να είναι $c_1 \approx c_2$.

Ο συντελεστής αδράνειας παίζει σημαντικό ρόλο στην ισορρόπηση της εξερεύνησης με την εκμετάλλευση (exploration-exploitation), καθώς με μεγάλη τιμή το βήμα αναζήτησης θα αυξηθεί πολύ και θα χαθούν τα τοπικά βέλτιστα. Αντίστοιχα, με μια μικρή τιμή θα περιοριστεί πολύ η περιοχή αναζήτησης. Η πιο συνηθισμένη τεχνική που χρησιμοποιείται για τον καθορισμό του συντελεστή αδράνειας είναι η γραμμική μείωση που ορίζεται ως εξής :

$$w = w_{max} - \frac{w_{max} - w_{min}}{T_{max}} * t \quad (2.3)$$

όπου t = αριθμός επανάληψης, T_{max} = μέγιστος αριθμός επαναλήψεων και συνήθως $w_{max} = 0.9$ και $w_{min} = 0.4$. Με αυτόν τον τρόπο ξεκινάει η αναζήτηση με μεγάλο βήμα και στη συνέχεια εξειδικεύεται σε μικρότερες περιοχές.

Οι τυχαίοι αριθμοί r_1, r_2 έχουν στοχαστικό χαρακτήρα και συμβάλουν στο να μην έχουμε γρήγορη σύγκλιση σε ένα P_{best} ενώ ταυτόχρονα να οδηγηθεί κάθε σωματίδιο στο G_{best} .

2.4.3 Βελτιστοποίηση Συμήνους Σωματιδίων στα Νευρωνικά Δίκτυα

Ο αλγόριθμος PSO είναι εύκολος στο να εφαρμοστεί καθώς το μαθηματικό του υπόβαθρο εξαρτάται κυρίως από τις εξισώσεις (2.1), (2.2) και χρησιμοποιείται με ιδιαίτερη επιτυχία στη ρύθμιση υπερπαραμέτρων των νευρωνικών δικτύων ώστε να επιτύχουν μεγαλύτερη ακρίβεια ή μικρότερο σφάλμα. [61]

Ιδιαίτερα ενδιαφέρον αποτελεί το γεγονός ότι πλέον ο ΒΣΣ φαίνεται να υπερισχύει της μεθόδου οπισθοδρόμησης (backpropagation) στην εκπαίδευση ενός νευρωνικού δικτύου με καθορισμένη αρχιτεκτονική αλλά με τα βάρη πλέον να παίρνουν τη θέση των μεταβλητών προς βελτιστοποίηση. [34, 60]

Το βασικό μειονέκτημα του ΒΣΣ είναι η ανάγκη του για υπολογιστική ισχύ καθώς για κάθε σωματίδιο απαιτείται πολλαπλός υπολογισμός της συνάρτησης εφαρμογής, και όσο μεγαλύτερος είναι ο αριθμός των σωματιδίων τόσο αυξάνεται η χρονική διάρκεια, η πολυπλοκότητα αλλά και η απαίτηση πόρων.

Κεφάλαιο 3

Ανάλυση Κατασκευής Μοντέλου

Στο κεφάλαιο αυτό, περιγράφεται η δομή και τα χαρακτηριστικά των μοντέλων που κατασκευάστηκαν.

Η πορεία αυτής της κατασκευής ξεκινά με τη συλλογή των δεδομένων, συνεχίζεται με την χρήση ενός Παφαγωγικού Αντιπαλικού Δικτύου όπως επίσης και με την τεχνική Επαύξησης Δεδομένων για την εξισορρόπηση της συλλογής μας. Έπειτα γίνεται χρήση τεχνικών Μεταφοράς Γνώσης σε προεκπαιδευμένα μοντέλα και αναζητούνται οι βέλτιστες υπερπαράμετροι με τον αλγόριθμο Βελτιστοποίησης Συμήνους Σωματιδίων.

Θα αναλυθούν ποιά δεδομένα χρησιμοποιήθηκαν, και ποια είναι η κατανομή τους. Επίσης όμως γίνεται επεξήγηση της αρχιτεκτονικής του ΠΑΔ όπως επίσης και ο τρόπος που εκπαιδεύτηκε. Θα γίνει αναφορά στα προεκπαιδευμένα μοντέλα που εφαρμόστηκαν και ποιες υπερπαράμετροι τέθηκαν σε αναζήτηση μέσω της ΒΣΣ.

3.1 Δεδομένα

Το πρώτο βήμα για την δημιουργία ενός μοντέλου που να μπορεί να ταξινομεί με μεγάλη ακρίβεια ακτινογραφίες πνευμόνων (X-rays) ασθενών που έχουν προσβληθεί από τον ιό SARS-CoV-2 ή covid19 (όπως όμως ονομαστεί η αντίστοιχη κλάση) ή είναι υγιείς, είναι η συλλογή δεδομένων.

Η συλλογή αυτών των δεδομένων ξεκίνησε με αναζήτηση στο διαδίκτυο για διαθέσιμες εικόνες και παρατηρήθηκε το γεγονός ότι υπήρχε ελάχιστο υλικό. Αυτό αποτέλεσε μια νέα πρόκληση, την δημιουργία ενός μοντέλου το οποίο να ανταποκρίνεται ακόμα και σε ένα μικρό διαθέσιμο σετ δεδομένων.

Το COVIDx Dataset, το σύνολο δεδομένων που δημιουργήθηκε και χρησιμοποιήθηκε, διαθέτει 18356 X-rays εικόνες χωρισμένες σε 3 κλάσεις (labeled

dataset). Πιο συγκεκριμένα, η κατανομή στις κλάσεις έχει ως εξής:

- Normal: 10192 55,5%
- Pneumonia: 7402 40,3%
- Covid19: 762 4,2%

Η συγκεκριμένη συλλογή δεδομένων προέκυψε μετά από την ενοποίηση και τον καθαρισμό των δεδομένων από διάφορα διαθέσιμα σύνολα δεδομένων από τις εξής πηγές:

- [12] Συλλογή από δημόσιες πηγές και από αρχείο νοσοκομείων
- [13] Συλλογή που έχει δημιουργηθεί για να ενισχύσει την εκπαίδευση των μοντέλων που αφορούν την αναγνώριση και την κατάταξη ρίσκου σε περιπτώσεις προσβολής από τον ιό.
- [14] Συλλογή που έχει δημιουργηθεί για να ενισχύσει την εκπαίδευση των μοντέλων που αφορούν αναγνώριση και την κατάταξη ρίσκου σε περιπτώσεις προσβολής από τον ιό.
- [15] Συλλογή για διαγωνισμό kaggle.
- [16] Συλλογή για διαγωνισμό kaggle αναγνώρισης πνευμονοπάθειας.

Στην συνέχεια έγινε ο κατάλληλος διαχωρισμός σε φακέλους Normal, Pneumonia, Covid-19 και ύστερα το σετ δεδομένων χωρίστηκε σε σετ εκπαίδευσης/επικύρωσης (train/validation) το οποίο θα ήταν αποκλειστικά για την εκπαίδευση (train) και την επικύρωση (validation) των μοντέλων και σε σετ ελέγχου το οποίο θα ήταν για τον τελικό έλεγχο του μοντέλου σε ένα τελείως ξένο για αυτό σετ δεδομένων. Ο χωρισμός έγινε ως εξής:

- Normal : 9292 για εκπαίδευση/επικύρωση 91,2% - 900 για τεστ 8,8%
- Pneumonia : 6602 για εκπαίδευση/επικύρωση 89,2% - 800 για τεστ 10,8%
- Covid19 : 682 για εκπαίδευση/επικύρωση 89,5% - 80 για τεστ 10,5%

Το σετ **εκπαίδευσης/επικύρωσης** έχει το **90,3%** του συνολικού σετ δεδομένων και **16576** εικόνες. Το **τεστ** σετ έχει το **9,7%** του συνολικού σετ δεδομένων και **1780** εικόνες.

Στο πλαίσιο αυτής της διπλωματικής, χρησιμοποιήθηκε η γλώσσα προγραμματισμού Python. Για την αποθήκευση των εικόνων, χρησιμοποιήθηκε ο

τύπος Numpy Array που είναι της μορφής [Τύπος, Πλάτος, Κανάλια Χρωμάτων] [17] με την τρίτη διάσταση πολλές φορές να παραλείπεται σε περιπτώσεις που παίρνει την τιμή 1. Η συγκεκριμένη συλλογή δεδομένων αποτελείται από εικόνες διαφορετικών μεγεθών και διαφορετικών καναλιών χρωμάτων. Αυτό αποτελεί πρόβλημα καθώς η δομή των μοντέλων βαθιάς μάθησης δημιουργείται για να εξυπηρετήσει δεδομένα ίδιων διαστάσεων και όχι πολλαπλών.

Επομένως έπρεπε να αντιμετωπιστεί το πρόβλημα των διαφορετικών διαστάσεων, όπως επίσης και το μη ισορροπημένο σετ δεδομένων, που προέρχεται από τις ελάχιστες εικόνες κλάσης covid-19, και είναι πιθανό να προκαλέσει υπερπροσαρμογή (overfitting).

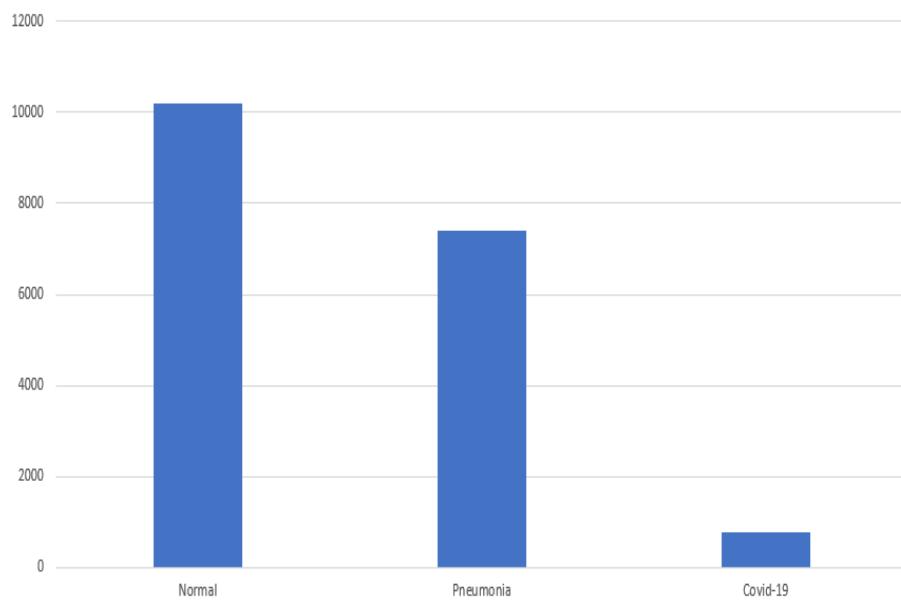


Figure 3.1: Κατανομή Δεδομένων

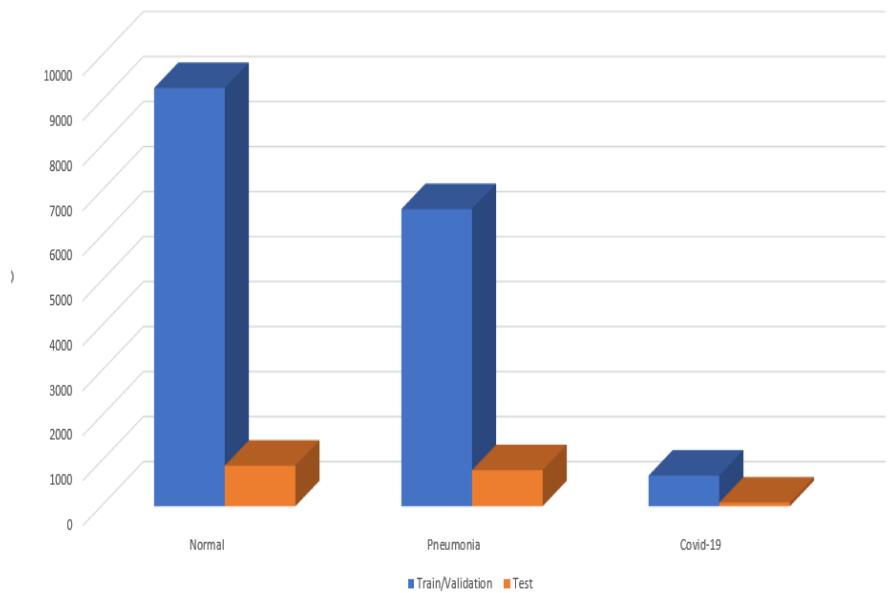


Figure 3.2: Κατανομή Δεδομένων



Figure 3.3: Covid-19



Figure 3.4: Normal



Figure 3.5: Pneumonia

3.2 Μοντέλο για Ισορρόπηση Δεδομένων

3.2.1 Παραγωγικό Αντιπαλικό Δίκτυο

Το πρόβλημα του μη ισορροπημένου σετ δεδομένων, έγινε η προσπάθεια να αντιμετωπιστεί και να εξομαλυνθεί μέσω της παραγωγής πλασματικών εικόνων που θα προέρχονται από ένα εκπαιδευμένο Παραγωγικό Αντιπαλικό Δίκτυο.

Το ΠΑΔ έχει τη δομή μοντέλου βαθιάς μάθησης, συνεπώς, όπως αναφέρθηκε, θα πρέπει να λαμβάνει σαν είσοδο, δεδομένα με τις ίδιες διαστάσεις. Για τον λόγο αυτό διαμορφώθηκαν και ανακατασκευάστηκαν οι εικόνες κατάλληλα σε ένα μέγευθος, με κριτήριο να παραχθεί όσο το δυνατόν καλύτερο οπτικό αποτέλεσμα αλλά και να έχει καλύτερη απόδοση το Παραγωγικό Αντιπαλικό Δίκτυο.

Όπως έχει αναφερθεί το Παραγωγικό Αντιπαλικό Δίκτυο αποτελείται από ένα Παραγωγικό και έναν Διαχωριστικό Δίκτυο, που στην συγκεκριμένη περίπτωση είναι είτε Συνελικτικά είτε Ανεστραμμένα Συνελικτικά Δίκτυα, όπου κάθε ένα από αυτά έχει τα αντίστοιχα βάρη που πρέπει να εκπαιδευθούν και να αναβαθμιστούν ανάλογα με την κατανομή των δεδομένων εισόδου. Μια εικόνα με υψηλή ανάλυση, π.χ. (1024,1024,3), απαιτεί μεγαλύτερη επεξεργασία για να μειωθούν οι διαστάσεις της, το οποίο οδηγεί σε περισσότερα βάρη που πρέπει να ενημερωθούν. Τα βάρη-μεταβλητές όσο περισσότερα είναι, όλο και μεγαλύτερο πλήθος δεδομένων θα χρειαστούν για να αποφευχθεί η περίπτωση της υποπροσαρμογής (underfit). Συνεπώς, πρέπει είτε να μειωθεί η ανάλυση της εικόνας είτε να μειωθούν τα στρώματα στο ΠΑΔ, με αρνητική συνέπεια, να συμβεί πιο απότομη μείωση διαστάσεων και πιθανή απώλεια στην πληροφορία χαρακτηριστικών των εικόνων.

Η ανάλυση η οποία αποφασίστηκε ήταν η **(64,64,1)** και η δομή του Παραγωγικού Δικτύου είναι η εξής:

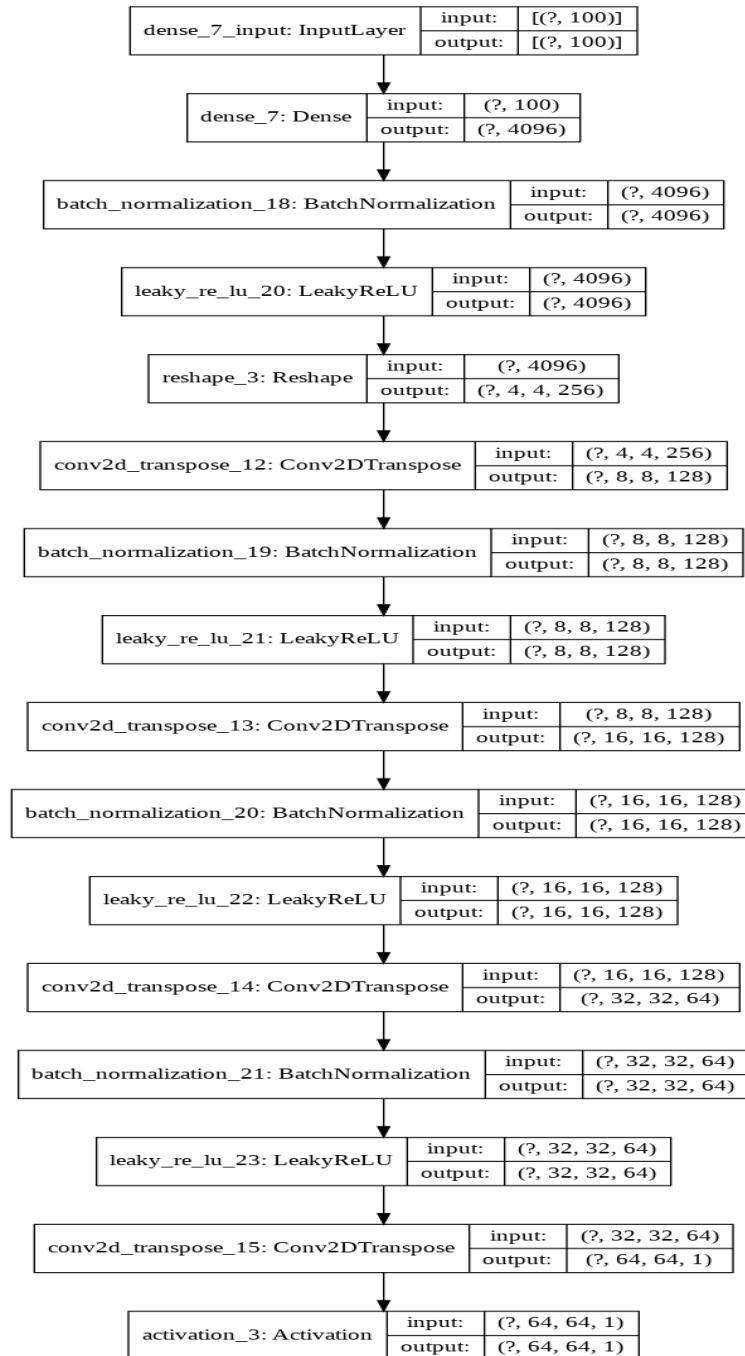


Figure 3.6: Δομή Παραγωγικού Δικτύου

Αρχικά χρησιμοποιήθηκε κανονική κατανομή για την αρχικοποίηση των βαρών. Έπειτα προστέθηκε ένα Πλήρως Συνδεδεμένο Στρώμα (Fully-Connected Layer) που δέχεται σαν είσοδο έναν τένσορα μεγέθους 100 που έχει οριστεί στον κώδικα, και παράγει έξοδο έναν τένσορα μεγέθους 4096 ($=4*4*256$). Ο τένσορας αυτός “περάστηκε” από Κανονικοποίηση Παρτιδών (Batch Normalization) και στρώμα “διαρροής” διορθωμένης γραμμικής μονάδας (Leaky Relu) ως στρώμα ενεργοποίησης. Τέλος με ένα Στρώμα Ανακατασκευής (Reshape Layer) έγινε λήψη του τένσορα μεγέθους (4,4,256), ο οποίος κατέληξε σε διαστάσεις (64,64,1).

Η βασική δομή για τον σκοπό αυτό είναι 3 μπλοκ το καθένα από :

- Ανεστραμμένο συνελικτικό στρώμα με βήμα = 2 για την σταδιακή αύξηση των διαστάσεων της εικόνας που οδηγούνται για ανακατασκευή.
- Κανονικοποίηση Παρτιδών (Batch Normalization) με momentum=0.9 το οποίο βοήθησε στην καλύτερη και πιο σταθερή εκπαίδευση όπως και στην αποφυγή του προβλήματος Βλάβης Μοντέλου που έχει αναλυθεί.
- Στρώμα ενεργοποίησης ”διαρροής” διορθωμένης γραμμικής μονάδας (LeakyReLU) με alpha = 0.1 για την αποφυγή του πρόβληματος ”νέκρωσης” νευρώνων (dying state), κατά το οποίο σταματάει η ενημέρωση των βαρών. [27]

Τέλος, τοποθετήθηκε ως συνάρτηση ενεργοποίησης η υπερβολική εφαπτομένη (tanh) για να επιτευχθεί σταθερή εκπαίδευση.

Η δομή του Διαχωριστικού Δικτύου είναι η εξής :

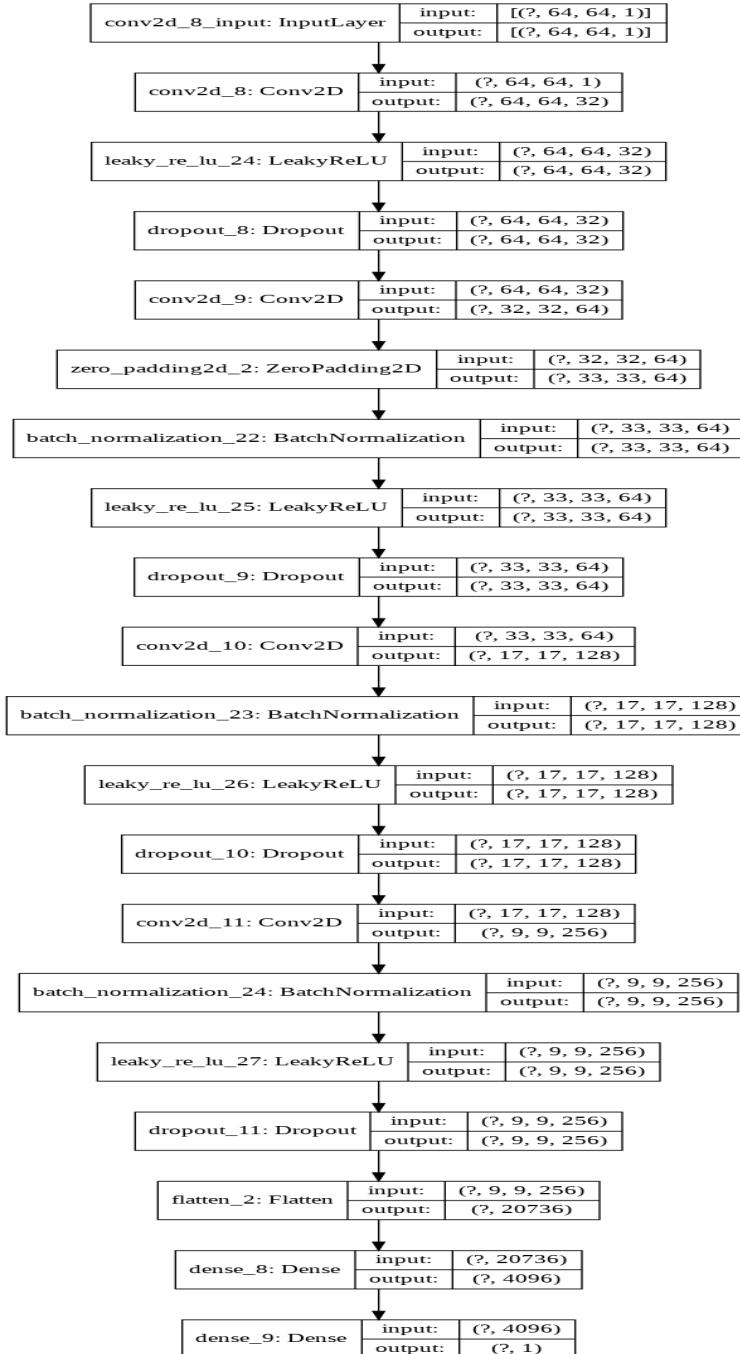


Figure 3.7: Δομή Διαχωριστικού Δικτύου

Ομοίως χρησιμοποιήθηκε κανονική κατανομή για την αρχικοποίηση των βαρών. Έστερα χρειάστηκε να μετατραπεί η είσοδος, που είναι ένας τένσορας (64,64,1) σε έναν τένσορα που είναι μονοδιάστατος, για να τοποθετηθεί σαν είσοδος σε σιγμοειδή συνάρτηση ενεργοποίησης και να ληφθεί η τελική απόφαση κατάταξης της εικόνας (0:πλασματική, 1:”αυθεντική”).

Η βασική δομή για το σκοπό αυτό είναι 4 μπλοκ το καθένα από :

- Συνελικτικό δίκτυο με μέγεθος πυρήνα (kernel size) = (4,4) και βήμα = 2 για τη σταδιακή μείωση διαστάσεων.
- Κανονικοποίηση Παρτιδών με momentum=0.9
- Συνάρτηση ενεργοποίησης LeakyReLU με alpha = 0.1
- Dropout Στρώμα με ρυθμό (rate) = 0.25 με στόχο την ενίσχυση της γενίκευσης

Τέλος, ύστερα από αυτήν τη διαδικασία λήφθηκε σαν αποτέλεσμα ένας τένσορας μεγέθους (9,9,256) ο οποίος μέσω ενός Επίπεδου (Flatten) Στρώματος στη σειρά με ένα Πλήρως Συνδεδεμένο Στρώμα (Fully-Connected Layer) μετατράπηκε σε μέγεθος (4096). Τέλος με ένα Πλήρως Συνδεδεμένο Στρώμα εξήχθη η τιμή που χρησιμοποιήθηκε στην σιγμοειδή συνάρτηση ενεργοποίησης. Ως optimizer χρησιμοποιήθηκε ο Adam με learning rate = 0.0002 και exponential decay rate = 0.5. Ο κώδικας για το ΠΑΔ που χρησιμοποιήθηκε, είναι διαθέσιμος στο [18].

Εφόσον ολοκληρώθηκε η αρχιτεκτονική του Παραγωγικού και Διαχωριστικού Δικτύου, τέθηκε σαν είσοδος του Διαχωριστικού η έξοδος του Παραγωγικού. Ορίστηκε ως συνάρτηση υπολογισμού σφάλματος, η **Δυαδική Εντροπία (Binary Cross Entropy)** και τέθηκε σαν optimizer, ο **Adam**. Το ΠΑΔ μετά από αυτές τις ρυθμίσεις ήταν ικανό να παράξει πλασματικές αλλά και αληθοφανείς εικόνες ασθενών με Covid-19.

3.2.2 Διαδικασία Εκπαίδευσης ΠΑΔ

Η εκπαίδευση του ΠΑΔ παρουσίασε πολλές αστάθειες και προβλήματα τα οποία ξεπεράστηκαν με ορισμένες τεχνικές, πολλές από αυτές τις αναφέρθηκαν στο Κεφάλαιο 2.1.3.

Έπειτα από πειράματα και δοκιμές (οι οποίες θα αναφερθούν στο Κεφάλαιο 4.1) προέκυψε το συμπέρασμα, να εκπαιδευτεί το μοντέλο για 1000 εποχές, καθώς παρατηρήθηκε ότι η ποιότητα της εικόνας μετά από αυτόν τον αριθμό, χαλάει αισθητά.

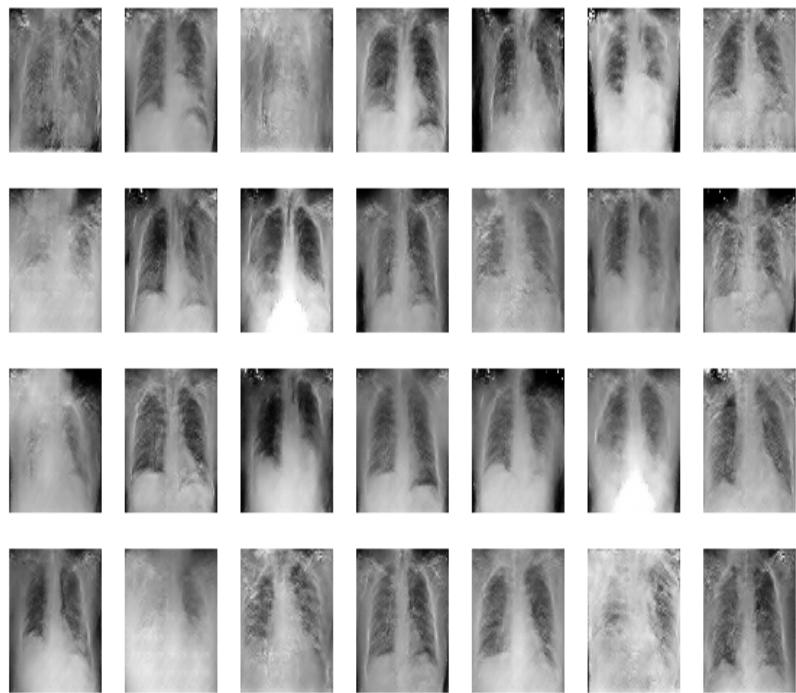


Figure 3.8: Αποτελέσματα ΠΑΔ

Παρατηρήθηκε μια "πτώση" στην ποιότητα της εικόνας συγχριτικά με τις εικόνες του αρχικού σετ δεδομένων, αλλά αυτό οφειλόταν χυρίως στο γεγονός ότι είχαν μειωθεί οι διαστάσεις τους και επομένως η ανάλυση τους. Ένα μικρό δείγμα των (64,64,1) διαστάσεων εικόνων που δέχθηκε το ΠΑΔ για την εκπαίδευση του παρατίθεται από κάτω.



Figure 3.9: Παράδειγμα Εισόδου - (64,64,1)



Figure 3.10: Παράδειγμα Εισόδου - (64,64,1)

Επιλέχθηκε η παρτίδα (batch) να ισούται με 32 για την εκπαίδευση, όπου η μισή σχηματίστηκε από "αυθεντικές" εικόνες και η άλλη μισή από πλασματικές εικόνες. Κατά τη διαμόρφωση της παρτίδας χρησιμοποιήθηκε :

- Θόρυβος (Noisy Labels): Δηλαδή αλλαγή του 5% της κλάσης των "αυθεντικών" εικόνων (από 1 σε 0) και 5% των πλασματικών εικόνων (από 0 σε 1)
- Εξομάλυνση Κλάσεων (Label Smoothing): Με αυτόν τον τρόπο πηγαίνει η κλάση 1 στο εύρος [0.7, 1.2] και η κλάση 0 στο εύρος [0.0, 0.3]

Τα παραπάνω πραγματοποιήθηκαν έτσι ώστε να επιτευχθεί μια πετυχημένη και ομαλή εκπαίδευση.

Τέλος, παρατηρήθηκε ότι το Παραγωγικό Δίκτυο παρήγαγε εικόνες τις οποίες το Διαχωριστικό Δίκτυο μπορούσε να ταξινομήσει σαν πλασματικές με αρκετά μεγάλη επιτυχία. Αυτό φάνηκε από την υψηλή ακρίβεια του Διαχωριστικού και από το μεγάλο σφάλμα του Παραγωγικού Δικτύου. Για να αντιμετωπιστεί αυτό το φαινόμενο διπλασιάστηκε η εκπαίδευση του Παραγωγικού Δικτύου έτσι ώστε να ενισχυθεί η προσαρμογή των βαρών του στην κατανομή των δεδομένων.

Αποθηκεύεται ανά 100 εποχές το Παραγωγικό Δίκτυο αλλά και οι εικόνες που παράγει. Με κριτήριο την αληθιοφάνεια των εικόνων αλλά και των διαγραμμάτων απώλειας και ακρίβειας του συνολικού ΠΑΔ έγινε η επιλογή του Παραγωγικού Δικτύου 1000ης εποχής το οποίο προσέθεσε 1500 νέες εικόνες στην κλάση covid19.

3.2.3 Επαύξηση Δεδομένων

Η παραπάνω διαδικασία οδήγησε σε αύξηση των εικόνων που ανήκουν στην κλάση covid19 με αποτέλεσμα να προκύψουν 2182 στοιχεία στο σετ εκπαίδευσης/επικύρωσης.

Επιχειρήθηκε μια ακόμα αύξηση με την τεχνική της Επαύξησης Δεδομένων. Πιο συγκεκριμένα, δύο θηράντες 2 πιθανές επιλογές :

- Οριζόντια Αντιστροφή
- Τυχαία Περιστροφή κατά (-25,25) μοίρες

Από αυτές τις δύο επιλογές έγινε μια τυχαία επιλογή για το ποια (μπορεί καμία ή μπορεί και οι δύο) θα εφαρμοστεί σε κάθε μια από τις εικόνες.

Με αυτόν τον τρόπο διπλασιάστηκαν οι εικόνες κλάσης covid19 και πλέον στο σετ εκπαίδευσης/επικύρωσης υπάρχουν 4364 στοιχεία. Η νέα κατανομή του σετ εκπαίδευσης/επικύρωσης προέκυψε ως εξής:

- Covid-19: 4364 - 21,6% από 682 - 4,2%
- Normal: 9292 - 45,9% από 9292 - 56%
- Pneumonia: 6602 - 32,5% από 6602 - 39,8%

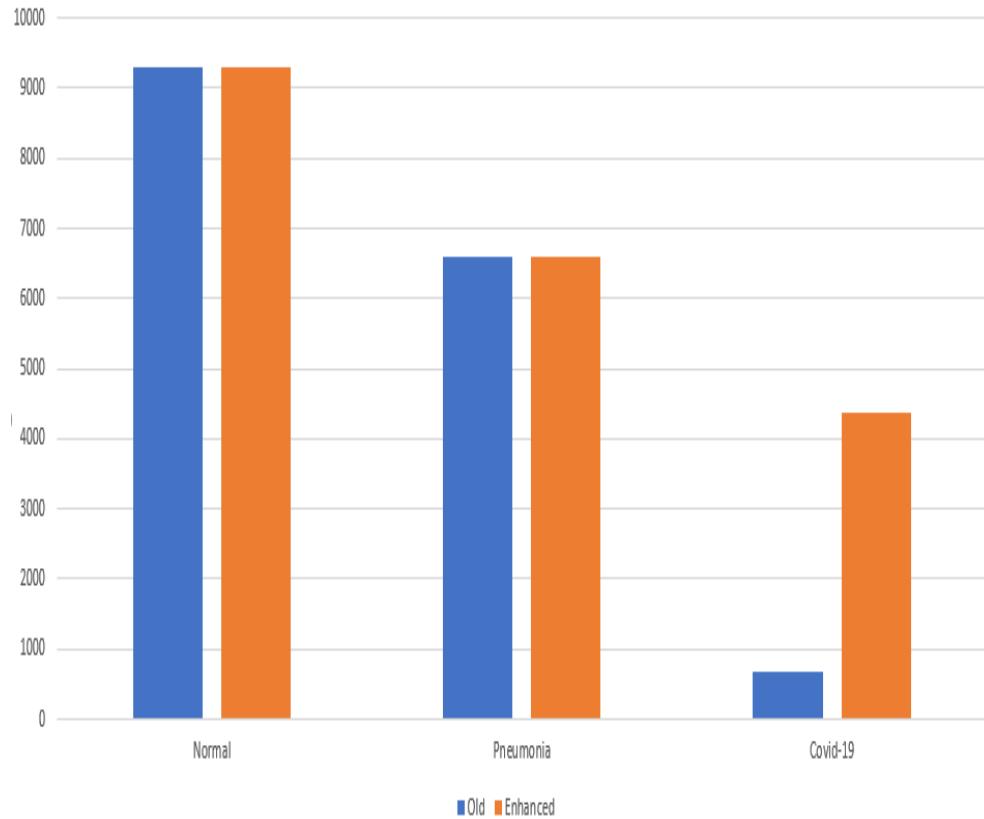


Figure 3.11: Train/Validation Distribution

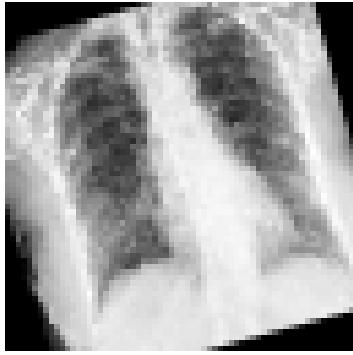


Figure 3.12: Τροποποιημένη Εικόνα 1

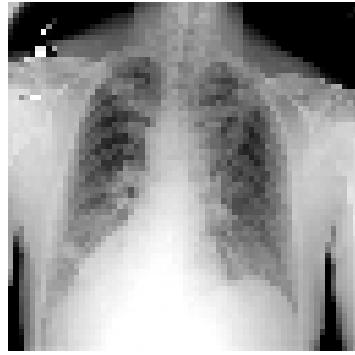


Figure 3.13: Τροποποιημένη Εικόνα 2

3.3 Επιλογή και Επεξεργασία Προεκπαιδευμένων Μοντέλων

Όπως αναφέρθηκε, σε περιπτώσεις με μικρό σετ δεδομένων, η τεχνική της Μεταφοράς Γνώσης και τα Προεκπαιδευμένα μοντέλα είναι πολύ χρήσιμα, καθώς διαθέτουν πληροφορίες και χαρακτηριστικά εικόνων.

Αρχικά, δουλεύοντας πάνω στο Keras framework, επιλέχθηκαν δύο προεκπαιδευμένα μοντέλα που παρέχονται εύκολα μέσω του API του. Πρώτη επιλογή ήταν το VGG19 [59] και η δεύτερη επιλογή ήταν το Resnet152V2 [36]. Τέλος χρησιμοποιήθηκε το μοντέλο CheXNet [19] το οποίο παρέχεται στο διαδίκτυο και έχει εκπαιδευτεί με 112,120 ακτινογραφίες πνευμόνων, οι οποίες έχουν πολλά κοινά χαρακτηριστικά με τις ακτινογραφίες του COVIDx Dataset.

Η επιλογή αυτών των μοντέλων έγινε με κριτήριο το πλήθος μεταβλητών-βαρών που έχουν προς εκπαίδευση. Πιο συγκεκριμένα κάθιε μοντέλο ανάλογα με το μέγεθος εισόδου έχει πλήθος μεταβλητών της τάξης :

- VGG19 : 150,000,000
- ResNet152V2 : 60,000,000
- CheXNet : Η βάση του είναι το DenseNet121 [37] οπότε 10,000,000 μεταβλητές.

Με αυτό τον τρόπο επιλέχθηκε ένα μοντέλο από κάθιε κλίμακα, έτσι ώστε να προκύψουν και συμπεράσματα για το μέγεθος του μοντέλου που θα ταίριαζε καλύτερα στο σετ αλλά και σε αντίστοιχες περιπτώσεις με περιορισμένο αριθμό δεδομένων.

3.3.1 VGG19

Στο Πανεπιστήμιο της Οξφόρδης ο Karen Simonyan, ο Andrew Zisserman και η ομάδα τους, η Visual Geometry Group (VGG), δημιούργησαν ένα βαθύ συνελικτικό νευρωνικό δίκτυο με 16-19 στρώματα, το VGG. Το μοντέλο αυτό συμμετείχε στον ετήσιο διαγωνισμό ImageNet Large-Scale Visual Recognition Challenge 2014 (ILSVRC) στον οποίο βγήκε 1o στην κατηγορία classification (κατηγοριοποίηση) και 2o στην κατηγορία localization (εντοπισμός). Έχει εκπαιδευτεί με πάνω από 1 εκατομμύριο εικόνες με αποτέλεσμα να έχει εξάγει πολλά χαρακτηριστικά. Η αρχιτεκτονική που ακολουθεί φαίνεται στην εικόνα 3.3.1. //

Το χαρακτηριστικό του μοντέλου είναι ότι:

- Τα Συνελικτικά Στρώματα που χρησιμοποιεί, έχουν όλα πυρήνα μεγέθους ($3x3$) με βήμα = 1 και γέμισμα (padding) = 1. Με αυτόν τον τρόπο διατηρεί σταθερές τις διαστάσεις του ύψους και πλάτους κάθε εικόνας ενώ αυξάνει τον αριθμό των φίλτρων.
- Τα Στρώματα Μέγιστης Συγκέντρωσης (Max Pooling Layers) έχουν φίλτρα μεγέθους ($2x2$) και βήμα = 1 με αποτέλεσμα να μειώνονται στο μισό οι διαστάσεις του ύψους και πλάτους και να παραμένει σταθερός ο αριθμός των φίλτρων.

Το VGG αποτελεί σημαντικό μοντέλο γιατί εισήγαγε την έννοια του βάθους για την εξαγωγή περισσότερων χαρακτηριστικών. Επίσης τονίζει την απλότητα σε κάθε του στρώμα καθώς μέχρι τότε υπήρχε η τάση για φίλτρα μεγάλων μεγεθών (π.χ. AlexNet μέγεθος φίλτρου = $(11x11)$).

Layer name	#Filters	#Parameters	#Activations
input			150K
conv1_1	64	1.7K	3.2M
conv1_2	64	36K	3.2M
max pooling			802K
conv2_1	128	73K	1.6M
conv2_2	128	147K	1.6M
max pooling			401K
conv3_1	256	300K	802K
conv3_2	256	600K	802K
conv3_3	256	600K	802K
conv3_4	256	600K	802K
max pooling			200K
conv4_1	512	1.1M	401K
conv4_2	512	2.3M	401K
conv4_3	512	2.3M	401K
conv4_4	512	2.3M	401K
max pooling			100K
conv5_1	512	2.3M	100K
conv5_2	512	2.3M	100K
conv5_3	512	2.3M	100K
conv5_4	512	2.3M	100K
max pooling			25K
fc6		103M	4K
fc7		17M	4K
output		4M	1K

Figure 3.14: VGG19 Architecture [30]

3.3.2 ResNet152V2

Το ResNet152 δημιουργήθηκε από τον Kaiming He και την ομάδα του και δημοσιεύθηκε στο άρθρο Deep Residual Learning for Image Recognition, το 2015. Το μοντέλο αυτό, κατάφερε να κερδίσει στον διαγωνισμό ILSVRC 2015 τόσο στη κατηγορία classification όσο και στη κατηγορία εντοπισμός και ταξινόμηση

αντικειμένων σε εικόνες όπως επίσης και στη κατηγορία εντοπισμού.

Όπως αναφέρθηκε, το 2014 με το VGG, εισήχθη η έννοια του μεγάλου βάθους στα δίκτυα. Αυτό όμως οδήγησε σε διάφορα προβλήματα, το βασικό ήταν το πρόβλημα της μηδενικής οπισθοδρόμησης (vanishing gradient), κατά το οποίο η τιμή που επιστρέφει στα πίσω στρώματα ο αλγόριθμος οπισθοδρόμησης (backpropagation) μηδενίζεται, με αποτέλεσμα να σταματάει η εκπαίδευση στα στρώματα που συμβάνει αυτό και να μην προχύπτει εξαγωγή χαρακτηριστικών στα πρώτα κυρίως στρώματα.

Για την επίλυση αυτού του προβλήματος εισήχθη η ιδέα του Μπλοκ ανατροφοδότησης (Residual Block). Το σκεπτικό πίσω από αυτό είναι ότι αν ένα πολύ βαθύ μοντέλο δεν μπορεί να έχει καλή απόδοση λόγω μηδενικής οπισθοδρόμησης θα μπορούσε να έχει τουλάχιστον την ίδια απόδοση με ένα πιο "ρηχό" μοντέλο.

Με την προσθήκη αυτού του τμήματος, η τιμή που επιστρέφει ο αλγόριθμος οπισθοδρόμησης έχει 2 διαδρομές να ακολουθήσει. Στην περίπτωση στρωμάτων που βρίσκονται προς το τέλος, η τιμή της παραγώγου που θα οδηγήσει στην ενημέρωση των βαρών δεν θα είναι 0 και συνεπώς η ευθεία πορεία θα ανανεώσει τα βάρη. Στα αρχικά στρώματα που η τιμή της παραγώγου έχει αρχίσει να μειώνεται, η πορεία μέσω της ταυτοτικής συνάρτησης (identity function) μπορεί να χρατήσει αλώβητη την τιμή της παραγώγου και επομένως τα αρχικά στρώματα να ανανεωθούν.

Στην 3.16 φαίνεται η δομή του μοντέλου σε συνδυασμό με την 3.17 που αναλύεται η εσωτερική δομή μεταξύ των Συνελικτικών Στρωμάτων (weight isovται με Convolutional)

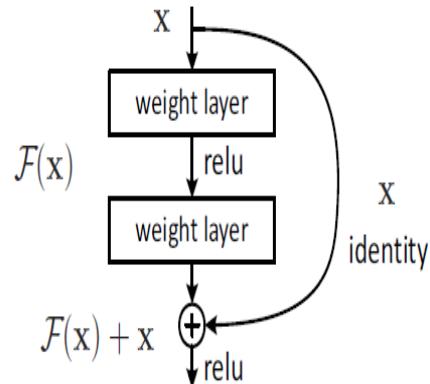


Figure 3.15: Τμήμα Ανατροφοδότησης [36]

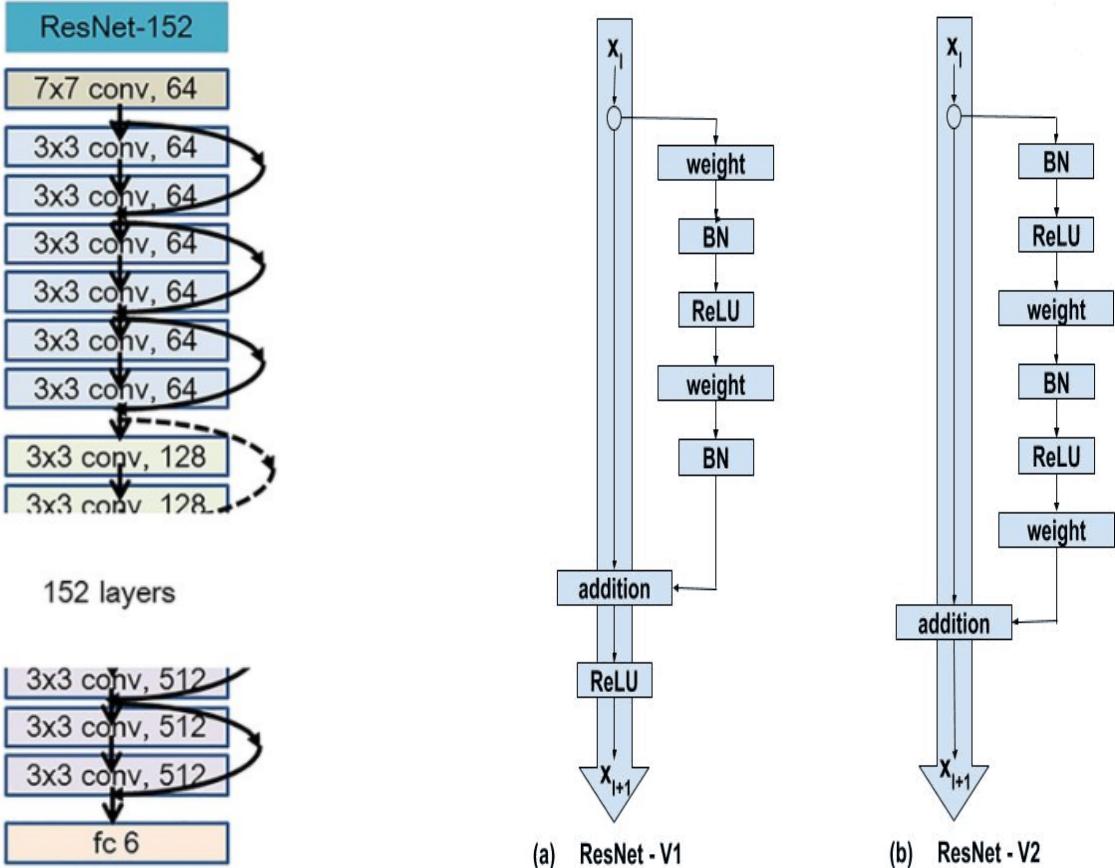


Figure 3.16: ResNet152 architecture[35]

Figure 3.17: ResNet152 vs ResNet152v2 [20]

3.3.3 CheXNet/Dense121

To CheXNet είναι ένα μοντέλο το οποίο δίνει τη δυνατότητα στο χρήστη να χρησιμοποιήσει κάποιο προεκπαιδευμένο μοντέλο μέσα από ένα πλήθος μοντέλων που προσφέρει. Επίσης, για κάθε μια από τις επιλογές προσφέρει και τα αντίστοιχα εκπαιδευμένα βάρη, όπως αυτά προέκυψαν από την εκπαίδευση πάνω σε 112,120 ακτινογραφίες πνευμόνων. Λόγω του πλήθους των παραμέτρων αλλά και για την αρχιτεκτονική του επιλέχθηκε στα πλαίσια αυτής της εργασίας το DenseNet121.

To DenseNet προτάθηκε το 2017 από τον Gao Huang και την ομάδα του στο άρθρο "Densely Connected Convolutional Networks" και απέσπασε το βραβείο Best Paper Award στο IEEE Συνέδριο με θέμα το Computer Vision and Pattern Recognition (CVPR). To DenseNet εισήγαγε μια νέα προσέγγιση για να λυθεί το

πρόβλημα της μηδενικής οπισθοδρόμησης.

Όπως αναλύθηκε και στο ResNet , η προσέγγιση που χρησιμοποιεί είναι το μονοπάτι συντόμευσης για να υπάρχει καλύτερη μεταφορά της τιμής της παραγώγου-σφάλματος, χωρίς να μειώνεται σε κάθε στρώμα που έχει βάρος. Έτσι, το DenseNet προτείνει κάθε στρώμα να δέχεται την πληροφορία από όλα τα προηγούμενα στρώματα, όπως φαίνεται στην 3.3.3. Σε αντίθεση με το ResNet που είχαμε άνθροιση, τα στρώματα στο DenseNet στοιβάζονται και πηγαίνουν ως πληροφορία στο επόμενο στρώμα. Κάθε στρώμα έχει ένα ρυθμό αύξησης (growth rate) το οποίο ορίζει τον αριθμό των στρωμάτων που θα έχει στην έξοδο του (στην περίπτωση του DenseNet121 παίρνει την τιμή 32).

Η στοιβαξη και μεταφορά των στρωμάτων γίνεται μεταξύ των DenseBlocks. Η δομή των DenseBlocks είναι η εξής:

- Στρώμα Κανονικοποίησης Παρτίδων
- ReLU συνάρτηση ενεργοποίησης
- Συνελικτικό στρώμα με μέγεθος φίλτρου = (1,1)
- Στρώμα Κανονικοποίησης Παρτίδων
- ReLU συνάρτηση ενεργοποίησης
- Συνελικτικό στρώμα με μέγεθος φίλτρου = (3,3)

Καθ'όλη τη διάρκεια ενός DenseBlock οι διαστάσεις (ύψος, πλάτος) παραμένουν σταθερές, συνεπώς πρέπει να υπάρχει κάποια μείωση σε αυτές ανάμεσα στα πολλαπλά DenseBlocks. Αυτό γίνεται με τα Μεταβατικά Στρώματα (Transition Layers) τα οποία έχουν την εξής δομή :

- Στρώμα Κανονικοποίησης Παρτίδων
- Συνελικτικό στρώμα με μέγεθος φίλτρου = (1,1)
- Στρώμα Μέσης Συγκέντρωσης (Averge Pooling Layer) διαστάσεων (2,2)

Το πως συνδυάζονται τα DenseBlocks και τα Μεταβατικά Στρώματα για μια ολοκληρωμένη δομή ενός DenseNet φαίνεται στην 3.3.3

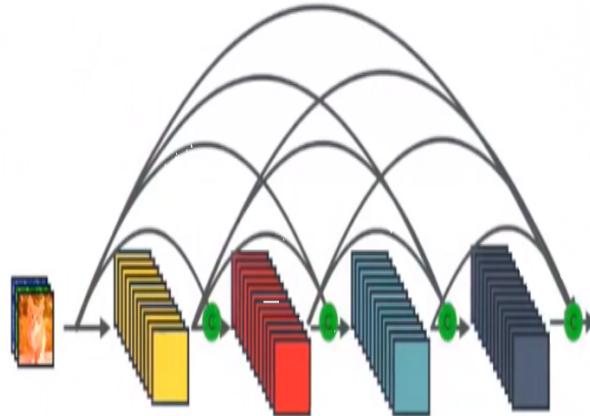


Figure 3.18: DenseBlocks [21]

Layers	Output Size	DenseNet-121	DenseNet-169	DenseNet-201	DenseNet-264
Convolution	112 × 112		7×7 conv, stride 2		
Pooling	56 × 56		3×3 max pool, stride 2		
Dense Block (1)	56 × 56	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$
Transition Layer (1)	56 × 56		1×1 conv		
	28 × 28		2×2 average pool, stride 2		
Dense Block (2)	28 × 28	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$
Transition Layer (2)	28 × 28		1×1 conv		
	14 × 14		2×2 average pool, stride 2		
Dense Block (3)	14 × 14	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 24$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 32$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 48$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 64$
Transition Layer (3)	14 × 14		1×1 conv		
	7 × 7		2×2 average pool, stride 2		
Dense Block (4)	7 × 7	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 16$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 32$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 32$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 48$
Classification Layer	1 × 1		7×7 global average pool		1000D fully-connected, softmax

Figure 3.19: DenseNets [22]

3.3.4 Ρύθμιση Στρωμάτων και Υπερπαραμέτρων

Έχοντας αναλύσει τα μοντέλα που θα εφαρμοστούν, εξηγείται ποια τεχνική ακολουθήθηκε στην εκπαίδευση και ποιες υπερπαραμετροί επιλέχθηκαν για ρύθμιση έτσι ώστε να ληφθεί η βέλτιστη απόδοση. Τα βήματα που τηρήθηκαν είναι τα εξής :

1. ”Φορτώνεται” κάθε μοντέλο. Το VGG19 και το ResNet152V2 είναι προεκπαίδευμένα με το ImageNet Dataset και το CheXnet είναι προεκπαίδευμένο με το σετ των ακτινογραφιών, που αναφέρθηκε.
2. Και τα τρία μοντέλα, υποχρεώνουν τη χρήση 3 καναλιών στις διαστάσεις των εικόνων, δηλαδή της μορφής (64,64,3). Παρ’ όλα αυτά, λόγω των απαιτήσεων του ΠΑΔ προκύπτουν εικόνες (64,64,1) , συνεπώς έγινε επανάληψη τριών φορών, των πρώτων δύο διαστάσεων για να δημιουργηθεί το επιθυμητό μέγεθος.
3. Δημιουργία δύο ειδών σετ δεδομένων: το πρώτο σετ είναι οι εικόνες χωρίς την προσθήκη των πλασματικών εικόνων, δηλαδή έχουν στην κλάση covid19 τις 682 ”αυθεντικές” εικόνες και το δεύτερο σετ είναι το ”ενισχυμένο” με τις 4364 εικόνες στην κλάση covid-19. Σε κάθε περίπτωση χρησιμοποιείται το 80% για εκπαίδευση και το 20% για επικύρωση .
4. Έγινε τροποποίηση κάθε μοντέλου μας με τις εξής υπερπαραμέτρους:

- ”Παγωμένα Στρώματα” (Freezing layers) : Στο VGG19 δόθηκε η επιλογή ο αριθμός των παγωμένων στρωμάτων να είναι [0,19]. Στο ResNet152V2 δόθηκε η επιλογή στο διάστημα [1,150] και στο CheXNet στο διάστημα [1,120].

Περίπτωση 1η

- Μετά από όλα τα Συνελικτικά στρώματα των μοντέλων, δόθηκε η δυνατότητα να εισαχθούν τα δεδομένα στα Πλήρη Συνδεδεμένα στρώματα είτε με ένα Επίπεδο (Flatten) στρώμα (Επιλογή 1) είτε με ένα Στρώμα Ολικής Μέσης Συγκέντρωσης (GlobalAveragePooling) (Επιλογή 0)
- Στη συνέχεια δόθηκε η επιλογή να μειωθούν οι διαστάσεις είτε με τρία Πλήρη Συνδεδεμένα στρώματα ($128 \rightarrow 64 \rightarrow 32$) (Επιλογή 3), είτε με 2 ($64 \rightarrow 32$) (Επιλογή 2), είτε με 1 (32) (Επιλογή 1). Όλα τα στρώματα έχουν συνάρτηση ενεργοποίησης την ReLU.

- Τέλος, προστέθηκε Dropout στρώμα μετά από κάθε Πλήρη Συνδεδεμένο στρώμα. Στην επιλογή με τα 3 Πλήρη Συνδεδεμένα στρώματα προκύπτουν 3 Dropout Layers, το 1ο έχει τιμή 0.5 το δεύτερο τοποθετείται σαν υπερπαράμετρος με πιθανές τιμές στο εύρος [0,0.65] ομοίως και για όλα τα υπόλοιπα στρώματα σε όλες τις άλλες επιλογές. Το τελευταίο Dropout στρώμα σε κάθε επιλογή είναι η υπερπαράμετρος νούμερο 5 και το προτελευταίο είναι η υπερπαράμετρος νούμερο 4.

Περίπτωση 2η

- Εφαρμόστηκε Επίπεδο (Flatten) στρώμα για την μετατροπή των δεδομένων σε μια διάσταση και έπειτα χρησιμοποιήθηκε ένα LSTM στρώμα με υπερπαράμετρο το πλήθος των μονάδων του (units) η οποία παίρνει τιμή [1,10] (υπερπαράμετρος νουμερο 2)
 - Χρησιμοποιήθηκε Dropout στρώμα (3η υπερπαράμετρος)
5. Προστέθηκε ένα Πλήρη Συνδεδεμένο στρώμα για την τελική ταξινόμηση, με συνάρτηση ενεργοποίησης την softmax, αφού πραγματεύεται 3 κλάσεις, και γίνεται compile με συνάρτηση υπολογισμού σφάλματος την Κατηγορηματική Εντροπία (Categorical Crossentropy) και optimizer = Adam.
 6. Εκπαιδεύτηκε κάθε μοντέλο μια φορά με το απλό σετ δεδομένων και μια φορά με το ενισχυμένο και στο τέλος πραγματοποιείται ένας τελικός έλεγχο με το σύνολο ελέγχου.
 7. Για την εκπαίδευση και για την εύρεση υπερπαραμέτρων χρησιμοποιήθηκε ο αλγόριθμος Βελτιστοποίησης Σμήνους Σωματιδίων που αναλύεται στην επόμενη ενότητα πώς εφαρμόστηκε.

3.3.5 Εφαρμογή Βελτιστοποίησης Σωματιδίων και Μετρικές

Αρχικά , για κάθε μοντέλο ορίστηκε η εκπαίδευση του που διαρκεί από 20 εποχές για το VGG19 και το ResNet152V2 και 50 εποχές για το CheXNet, με μέγεθος παρτίδας (batch) = 1000.

Για την εφαρμογή της ΒΣΣ δημιουργήθηκαν δύο πίνακες , ο πρώτος διαθέτει τα κάτω αριθμητικά όρια και ο δεύτερος τα πάνω αριθμητικά όρια κάθε υπερπαραμέτρου.

Οι τιμές των μεταβλητών της ΒΣΣ που χρησιμοποιήθηκαν είναι :

- 10 σωματίδια που αναζητούν την βέλτιστη θέση
- 30 αναζητήσεις-μετακινήσεις των particles (τερματική συνθήκη)
- συντελεστές επιτάχυνσης $c1, c2 = 0.5$

Η ΒΣΣ ψάχνει να βρει την καλύτερη τιμή για την μετρική: ακρίβεια (accuracy). Παρόλα αυτά εξετάστηκαν και άλλες μετρικές. Πιο συγκεκριμένα γίνεται αναφορά σε :

- Ευστοχία (Presicion) X% : Όσες φορές η κλάση A βγήκε σαν πρόβλεψη , X% ήταν σωστή η πρόβλεψη

- Ευαισθησία (Recall) $\Upsilon\%$: Όσες φορές η κλάση A θα έπρεπε να βγει σαν πρόβλεψη, $\Upsilon\%$ βγήκε.
- F1-score : $(2 * Precision * Recall) / (Precision + Recall)$ δίνει ένα συνδυασμό των προηγούμενων δύο μετρικών.
- Micro Μέσος : Υπολογίζει τον μέσο όρο κάθε μετρικής με συντελεστή 1 για κάθε κλάση
- Σταθμισμένος Μέσος (Weighted Average) : Υπολογίζει τον μέσο όρο κάθε μετρικής , αλλά ο συντελεστής κάθε κλάσης είναι το ποσοστό των δειγμάτων της ως προς το συνολικό πλήθος του σετ

Κεφάλαιο 4

Πειραματικά Δεδομένα

4.1 Παραγωγικό Αντιπαλικό Δίκτυο

Στο προηγούμενο Κεφάλαιο ορίστηκε η τελική δομή η οποία χρησιμοποιήθηκε για την ενίσχυση του σετ δεδομένων με περαιτέρω πλασματικές εικόνες. Στην ενότητα αυτή θα παρουσιαστεί η διαδικασία επιλογής της συγκεκριμένης αρχιτεκτονικής, παραθέτοντας τα αποτελέσματα εικόνων και διαγραμμάτων της απόδοσης του Παραγωγικού και Διαχωριστικού Δικτύου, ανά εποχή εκπαίδευσης.

Σημειώνεται πως τα βασικά κριτήρια επιλογής του τελικού μοντέλου ήταν η σύγκλιση των τιμών του σφάλματος (loss) σε ένα εύρος [0.3,0.5] για το Διαχωριστικό Δίκτυο και [1,2] για το Παραγωγικό Δίκτυο, όπως επίσης και το οπτικό αποτέλεσμα, δηλαδή η ποιότητα και η αληθιοφάνεια της παραγόμενης εικόνας. Στις περιπτώσεις που χρησιμοποιήθηκε εξομάλυνση κλάσεων (label smoothing), δεν ήταν δυνατή η παρουσίαση διαγράμματος ακρίβειας καθώς θα ήταν μηδενικό.

4.1.1 Εικόνες Μεγέθους (128,128,1)

Αρχικά οι πρώτες προσπάθειες παραγωγής εικόνων ήταν για εικόνες καλύτερης ανάλυσης ή μεγαλύτερων διαστάσεων, δηλαδή (256,256,3). Παρατηρήθηκε όμως ότι τα οπτικά αποτελέσματα δεν ήταν ικανοποιητικά καθώς και το σφάλμα του Παραγωγικού Δικτύου αδύνατούσε να συγκλίνει σε μια τιμή. Παρακάτω παρατίθενται τα αποτελέσματα για το ΠΑΔ μοντέλο εικόνων με ανάλυση (128,128,1).

.4.1.1 4.1.1 4.1.1 4.1.1 4.1.1

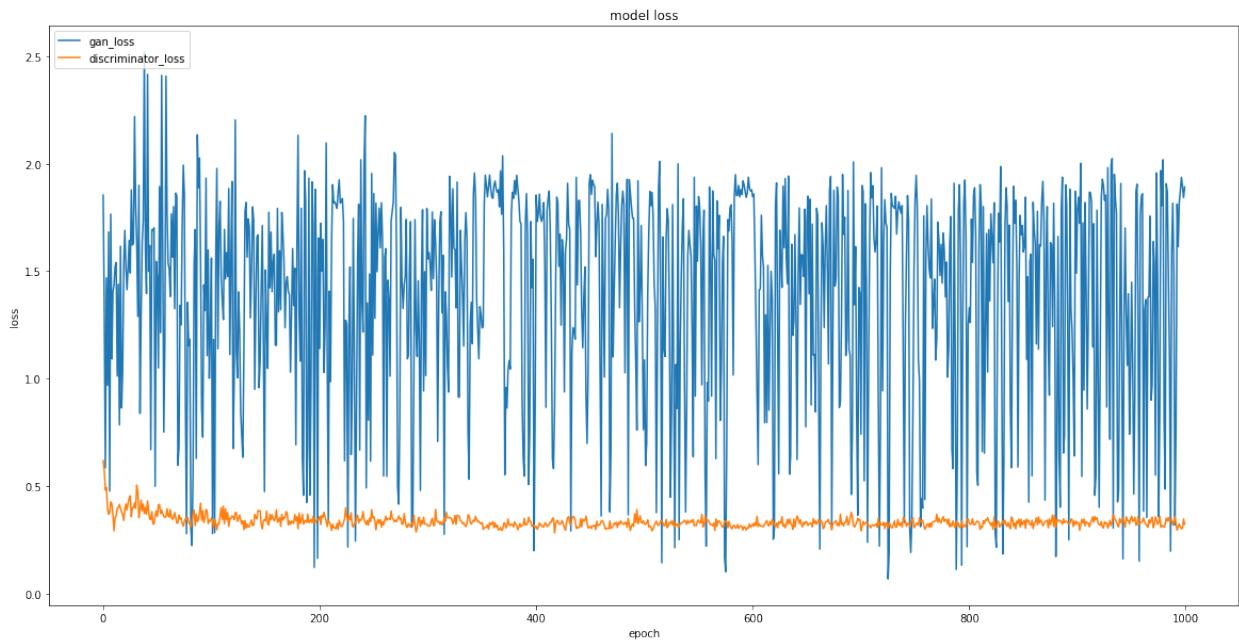


Figure 4.1: Διάγραμμα σφάλματος για (128,128,1) εικόνες

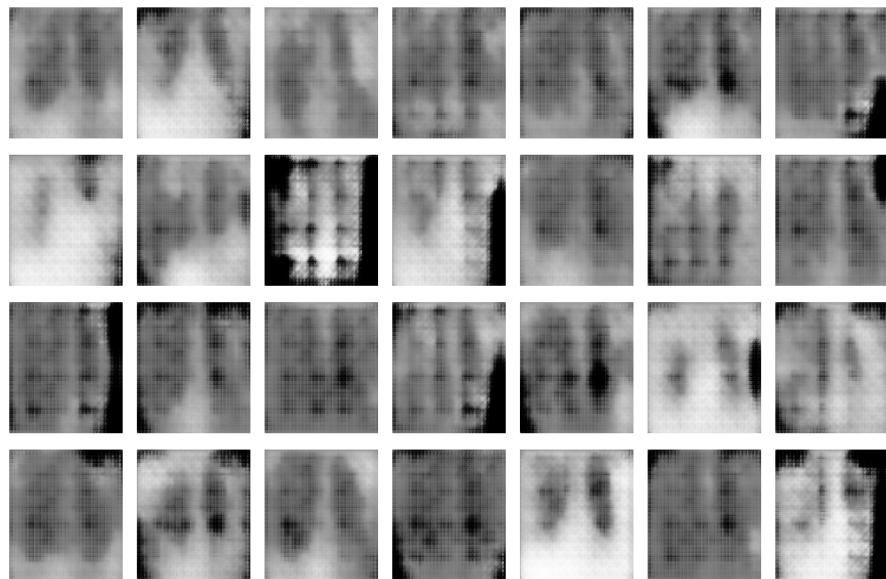


Figure 4.2: 200 Εποχές (128,128,1)

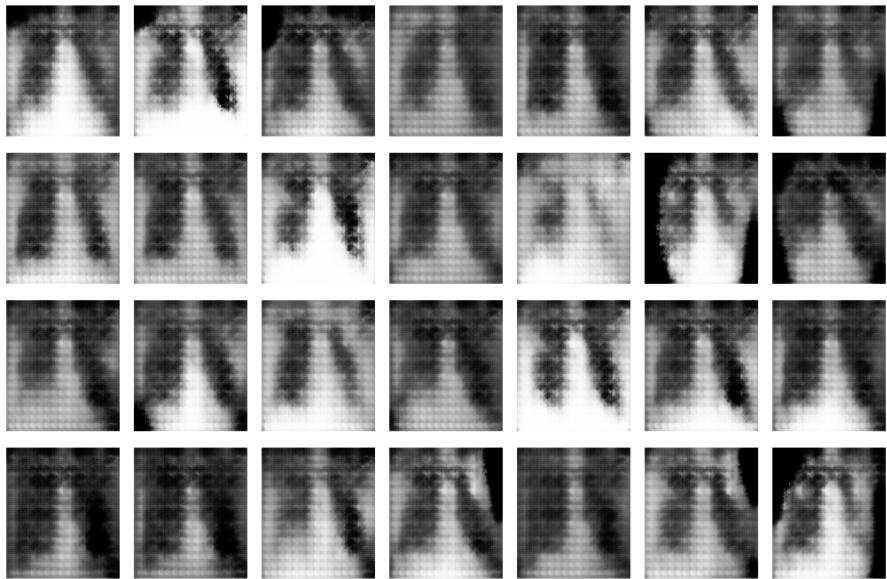


Figure 4.3: 500 Εποχές (128,128,1)



Figure 4.4: 800 Εποχές (128,128,1)

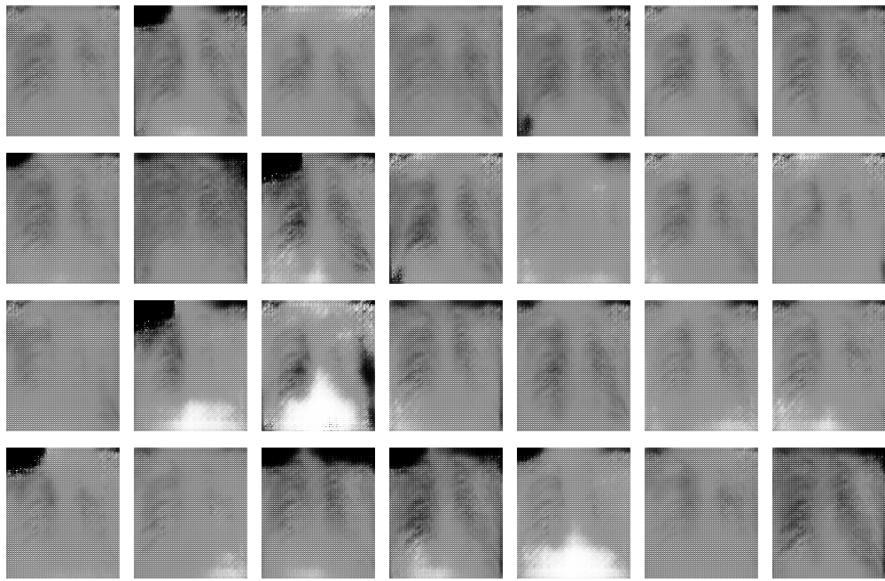


Figure 4.5: 1000 Εποχές (128,128,1)

Στα παραπάνω παρατηρήθηκε το φαινόμενο της Βλάβης Μοντέλου (2.1.3). Αυτό είναι εμφανές από το οπτικό αποτέλεσμα καθώς οι εικόνες έτειναν να μοιάσουν όλο και περισσότερο η μία στην άλλη ενώ ”θολώνουν” τα χαρακτηριστικά τους. Η διαπίστωση αυτή ενισχύθηκε και από το διάγραμμα σφάλματος αφού το σφάλμα του Παραγωγικού Δικτύου εμφάνισε ταλάντωση, αποτέλεσμα των εναλλαγών των καταστάσεών του.

4.1.2 Εικόνες Μεγέθους (64,64,3)

Στη συνέχεια πραγματοποιήθηκε μια ακόμα αλλαγή στις διαστάσεις των εικόνων και δοκιμάστηκε ανάλυση (64,64,3) με τα αποτελέσματα να παρατίθενται παρακάτω.

4.1.2 4.1.2 4.1.2 4.1.2 4.1.2

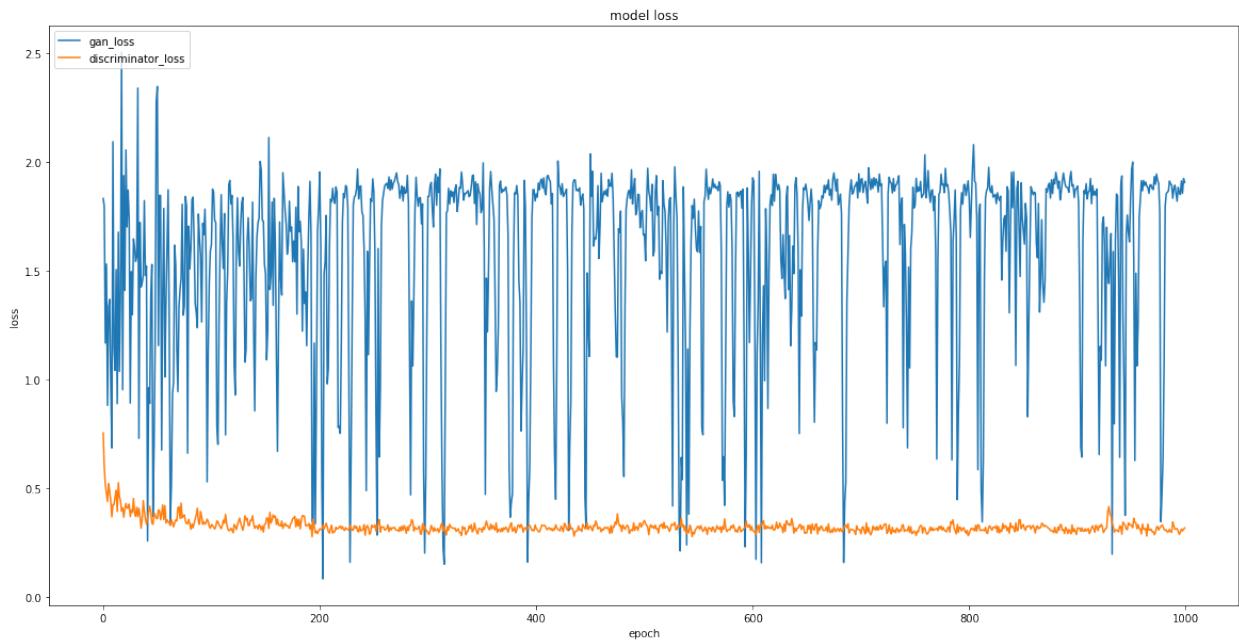


Figure 4.6: Διάγραμμα σφάλματος για (64,64,3) εικόνες

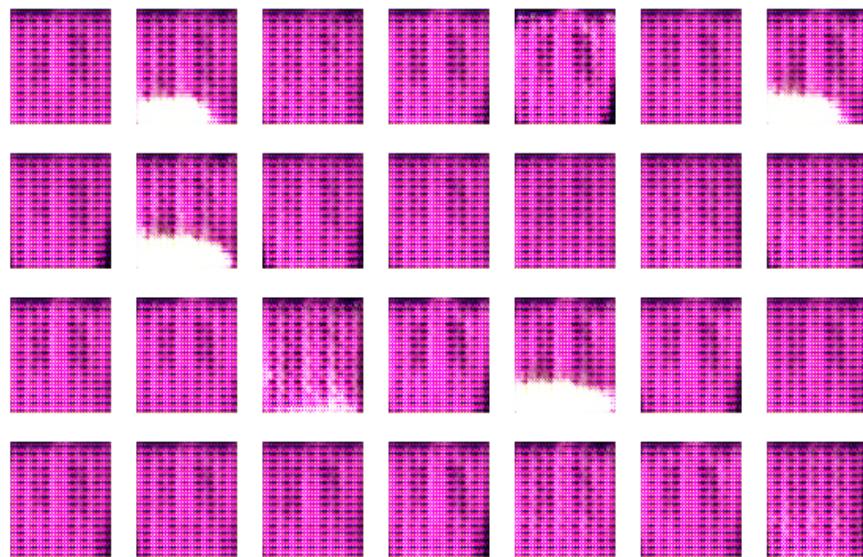


Figure 4.7: 200 Εποχές (64,64,3)



Figure 4.8: 500 Εποχές (64,64,3)

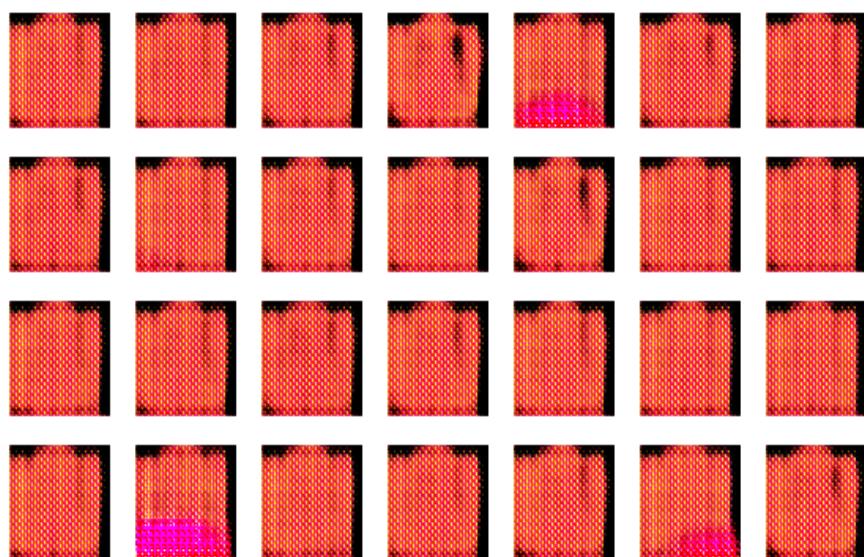


Figure 4.9: 800 Εποχές (64,64,3)

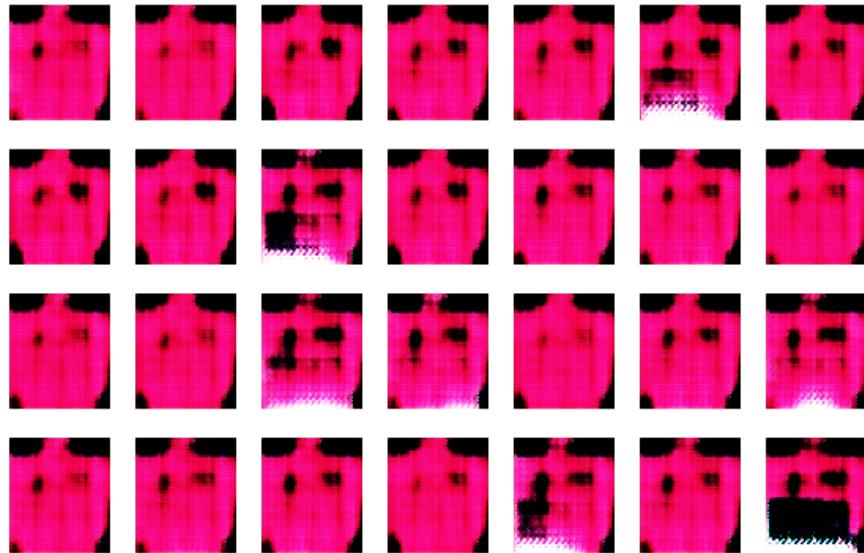


Figure 4.10: 1000 Εποχές (64,64,3)

Για ακόμα μια φορά προέκυψε το φαινόμενο της Βλάβης Μοντέλου. Φαίνεται ξεκάθαρα στις 1000 εποχές οι εικόνες να μοιάζουν όλο και περισσότερο μεταξύ τους, όπως επίσης παρατηρείται ταλάντωση στο σφάλμα του Παραγωγικού Δικτύου.

4.1.3 Εικόνες Μεγέθους (64,64,1)

Ακολούθως πραγματοποιήθηκε άλλη μια τροποποίηση στις διαστάσεις και πιο συγκεκριμένα έγινε μείωση σε (64,64,1) με τα αποτελέσματα να εμφανίζονται παρακάτω. 4.1.3 4.1.3 4.1.3 4.1.3 4.1.3

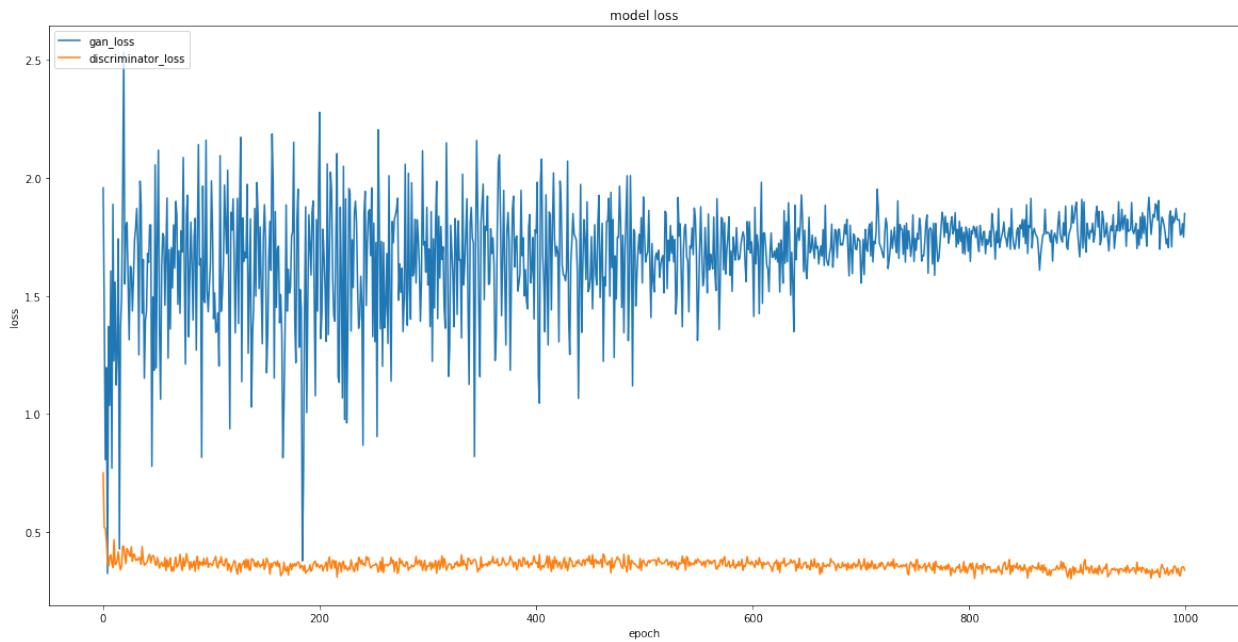


Figure 4.11: Διάγραμμα σφάλματος για $(64,64,1)$ εικόνες

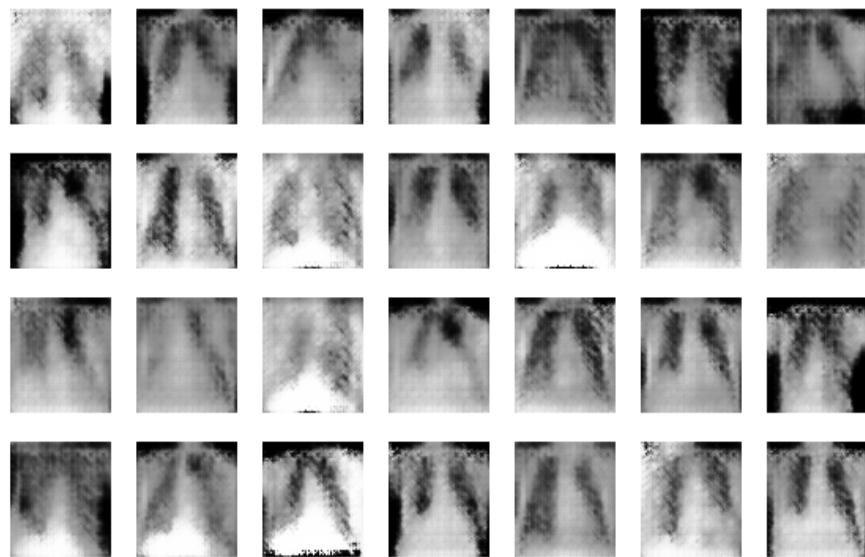


Figure 4.12: 200 Εποχές $(64,64,1)$

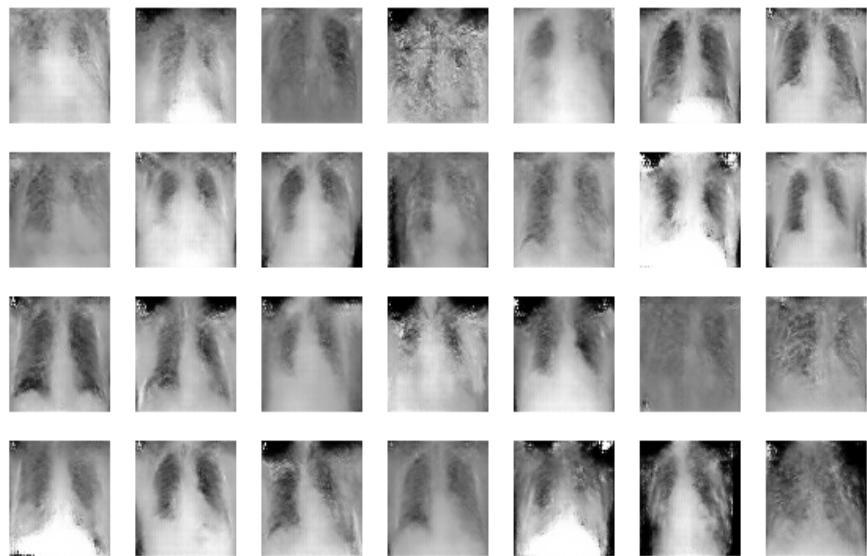


Figure 4.13: 500 Εποχές (64,64,1)

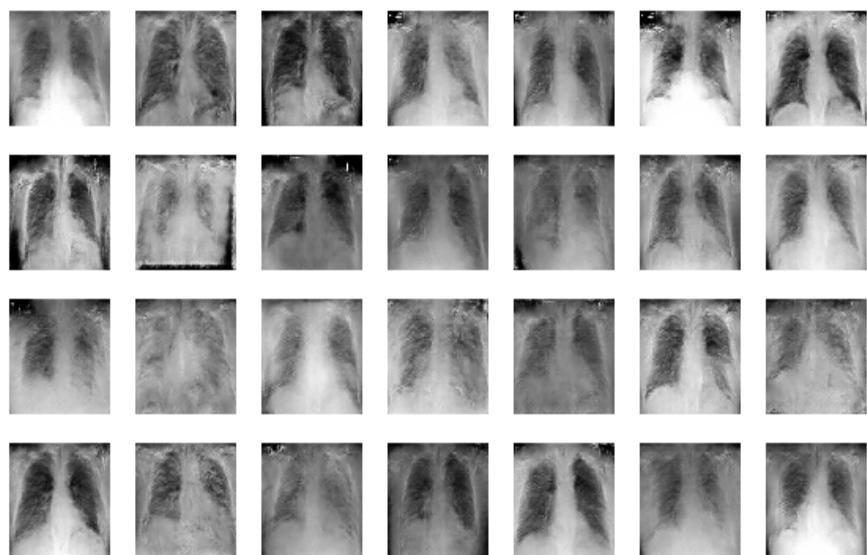


Figure 4.14: 800 Εποχές (64,64,1)

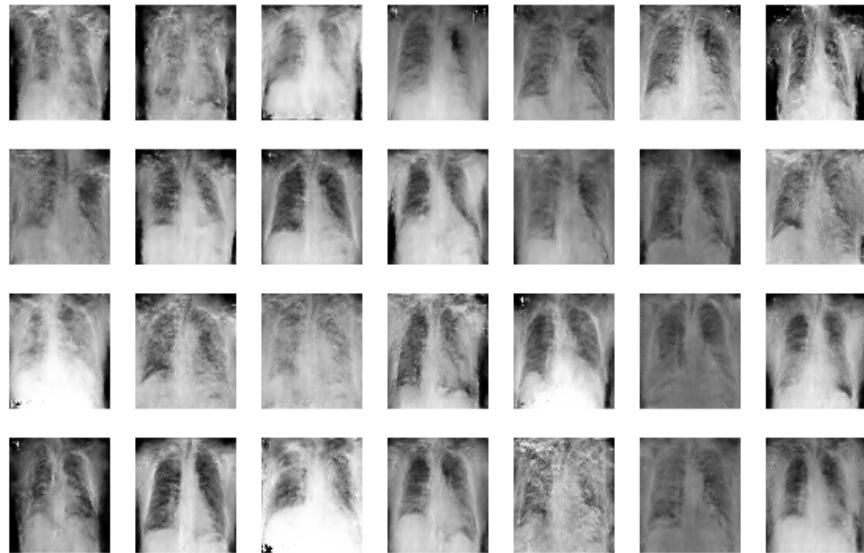


Figure 4.15: 1000 Εποχές (64,64,1)

Παρατηρήθηκε πως το οπτικό αποτέλεσμα ήταν αρκετά ικανοποιητικό, λαμβάνοντας υπόψιν και την ποιότητα της εικόνας που χρησιμοποιήθηκε ως είσοδος (βλ. προηγούμενη ενότητα 3.2.2). Επίσης παρατηρήθηκε για πρώτη φορά σύγκλιση στο Παραγωγικό Δίκτυο καθώς επίσης το Διαχωριστικό Δίκτυο κυμάνθηκε σε αποδεκτά πλαίσια γύρω στο 35%-40%. Συνεπώς, αυτή ήταν και η τελική επιλογή που προτιμήθηκε. Στα επόμενα πειράματα όμως αναδειχθεί η επιρροή στην εκπαίδευση αυτού του μοντέλου αν επιλεχθεί μια διαφορετική αρχιτεκτονική ή ακόμα και αν δεν χρησιμοποιηθούν τεχνικές για σταθερότητα.

4.1.4 Μοντέλο χωρίς Θόρυβο και Εξομάλυνση Κλάσεων

Σε αυτό το πείραμα αφαιρέθηκε ο επιπρόσθετος θόρυβος, όπως και η τεχνική εξομάλυνσης κλάσεων, διατηρώντας την ίδια αρχιτεκτονική με το μοντέλο που επιλέχθηκε με ανάλυση εικόνας (64,64,1). Ακολουθούν τα αποτελέσματα, όπως και το διάγραμμα ακρίβειας του μοντέλου, το οποίο παρείχε χρήσιμη πληροφορία για το λόγο για τον οποίο προτιμήθηκε το συγκεκριμένο μοντέλο. 4.1.4 4.1.4 4.1.4 4.1.4 4.1.4 4.1.4.

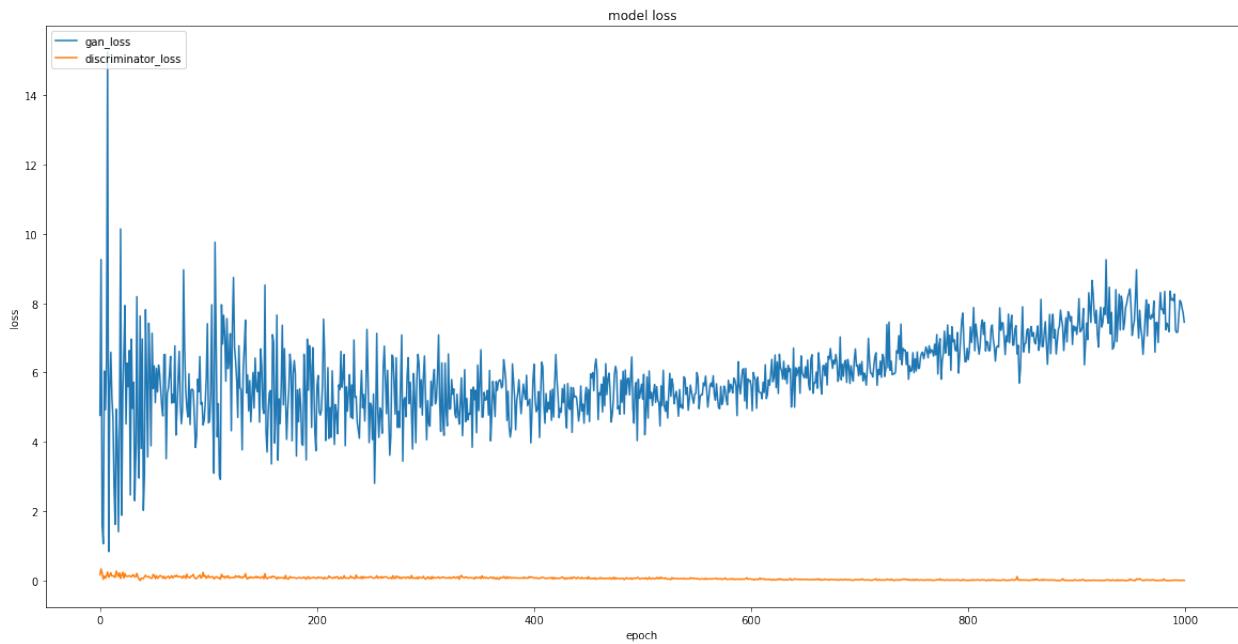


Figure 4.16: Διάγραμμα σφάλματος για (64,64,1) εικόνες - Ασταθές

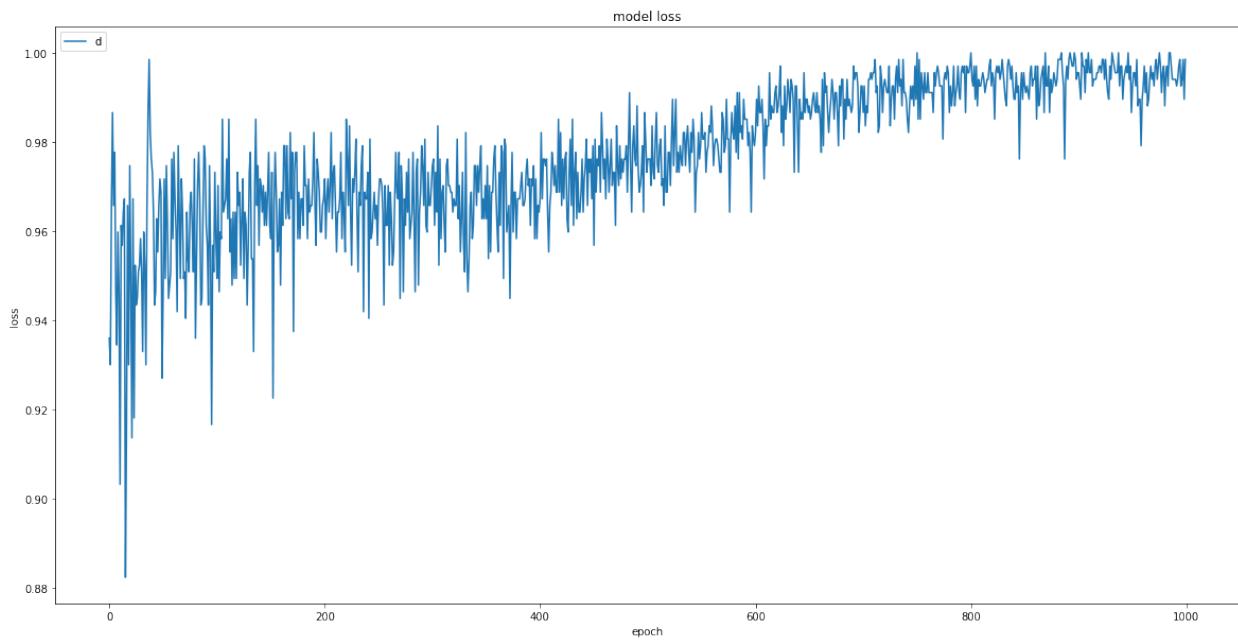


Figure 4.17: Διάγραμμα ακρίβειας για (64,64,1) εικόνες - Ασταθές

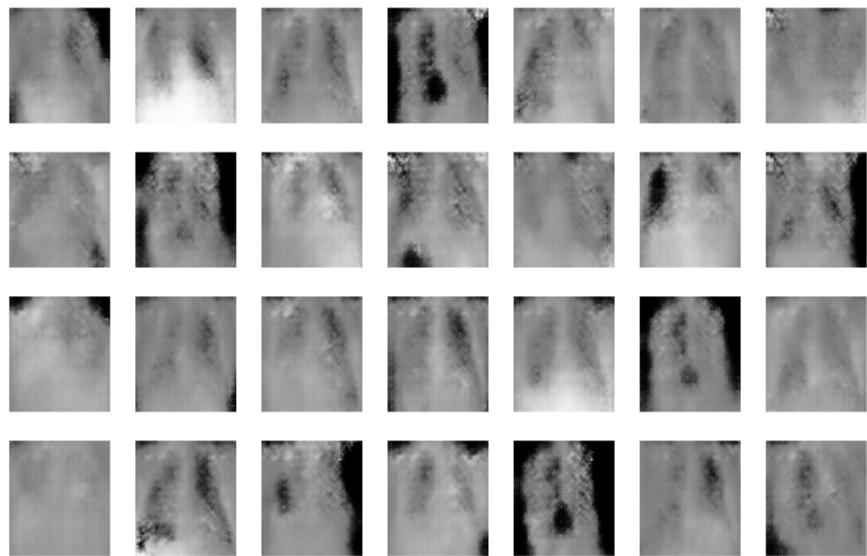


Figure 4.18: 200 Εποχές (64,64,1) - Ασταθές

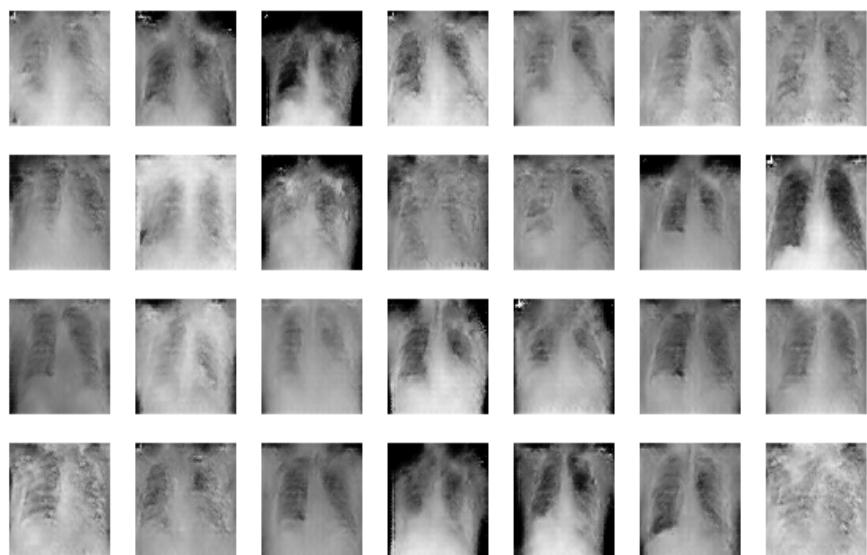


Figure 4.19: 500 Εποχές (64,64,1) - Ασταθές

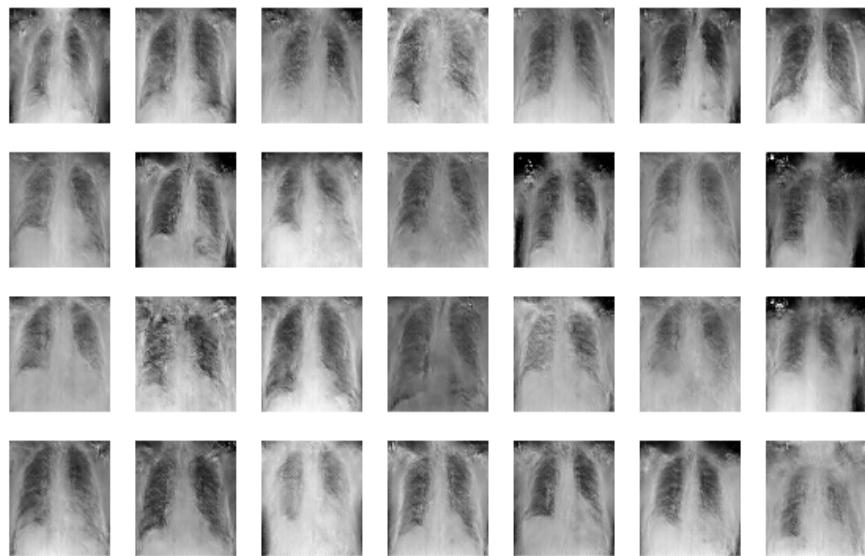


Figure 4.20: 800 Εποχές (64,64,1) - Ασταθές

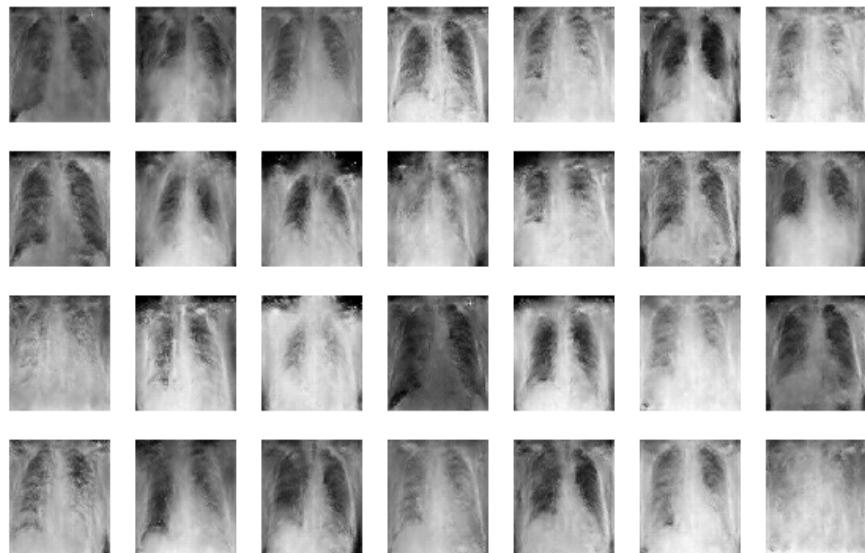


Figure 4.21: 1000 Εποχές (64,64,1) - Ασταθές

Αρχικά παρατηρήθηκε ότι τα αποτελέσματα των εικόνων δεν απείχαν ιδιαίτερα σε ποιότητα από την προηγούμενη περίπτωση, ωστόσο φάνηκε πόσο "αδύναμο" ήταν το

Παραγωγικό Δίκτυο σε σχέση με το Διαχωριστικό. Αυτό γίνεται αντιληπτό στο διάγραμμα σφάλματος (4.16) όπου το σφάλμα του Διαχωριστικού Δίκτυου προσεγγίζει μηδενικές τιμές και το σφάλμα του ΠΑΔ έχει μεγαλώσει αρκετά, καθώς και στο διάγραμμα ακρίβειας (4.17). Το Διαχωριστικό Δίκτυο συγκλίνει στο 100%, δηλαδή το Παραγωγικό Δίκτυο δεν καταφέρνει να είναι αρκετά "πειστικό" με τις εικόνες που παρήγαγε και οδηγώντας σε πρόβλημα Αποτυχίας Σύγκλισης.

4.1.5 Μοντέλο με μικρότερο αριθμό στρωμάτων

Στο πείραμα αυτό μειώθηκαν κατά ένα στρώμα το Παραγωγικό και Διαχωριστικό Δίκτυο, συγκεκριμένα τα στρώματα που αφορούν την αύξηση των διαστάσεων (Transpose Convolutional) στο Παραγωγικό Δίκτυο και τα στρώματα που αφορούν στη μείωση των διαστάσεων (Convolutional) στο Διαχωριστικό Δίκτυο.

Πιο συγκεκριμένα οι παράμετροι στο Παραγωγικό Δίκτυο άλλαξαν ως εξής:

- Total params: 1,119,841 → 1,021,217
- Trainable params: 1,111,201 → 1,012,705
- Non-trainable params: 8,640 → 8,512

Στο Διαχωριστικό Δίκτυο η αντίστοιχη αλλαγή ήταν η εξής:

- Total params: 26,913,505 → 17,377,697
- Trainable params: 26,912,609 → 17,376,929
- Non-trainable params: 896 → 768

Η μείωση των διαστάσεων έγινε πιο απότομα, με αποτέλεσμα το μοντέλο να συλλέξει λιγότερες πληροφορίες για την κατανομή των δεδομένων. Παρατίθενται τα αποτελέσματα. 4.1.5 4.1.5 4.1.5 4.1.5 4.1.5.

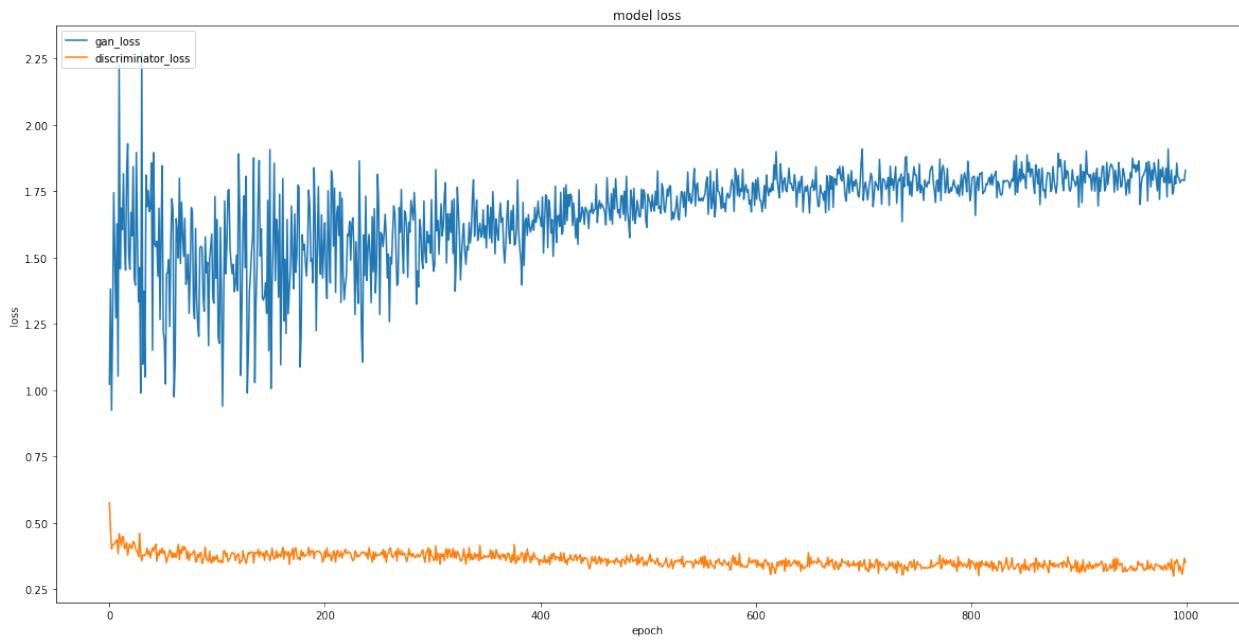


Figure 4.22: Διάγραμμα σφάλματος για (64,64,1) εικόνες - Μικρότερο Μοντέλο

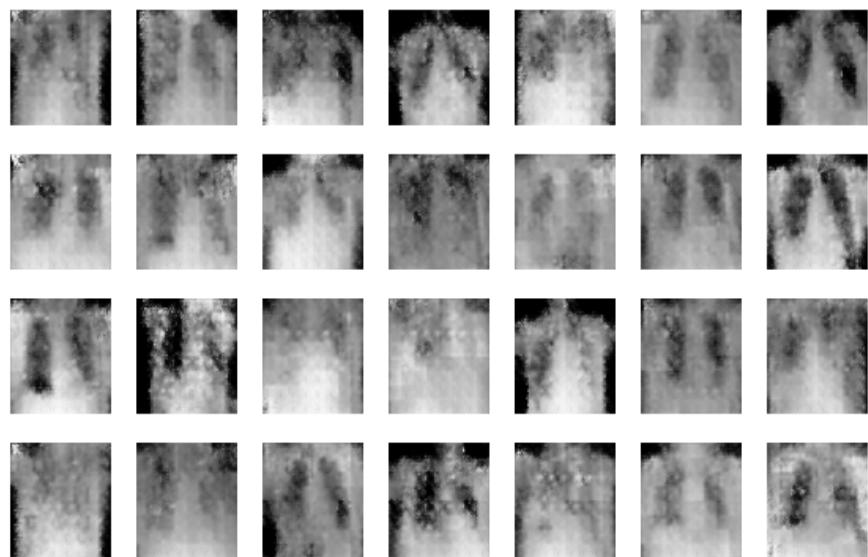


Figure 4.23: 200 Εποχές (64,64,1) - Μικρότερο Μοντέλο

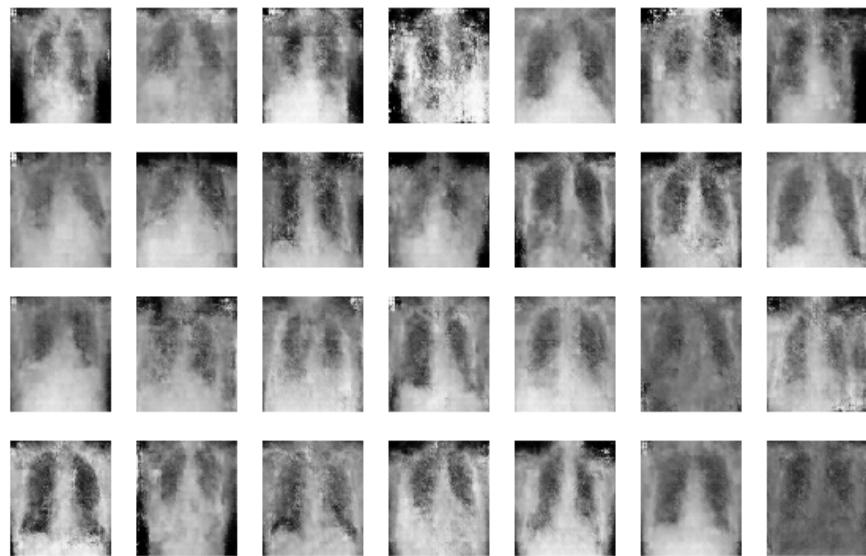


Figure 4.24: 500 Εποχές (64,64,1) - Μικρότερο Μοντέλο

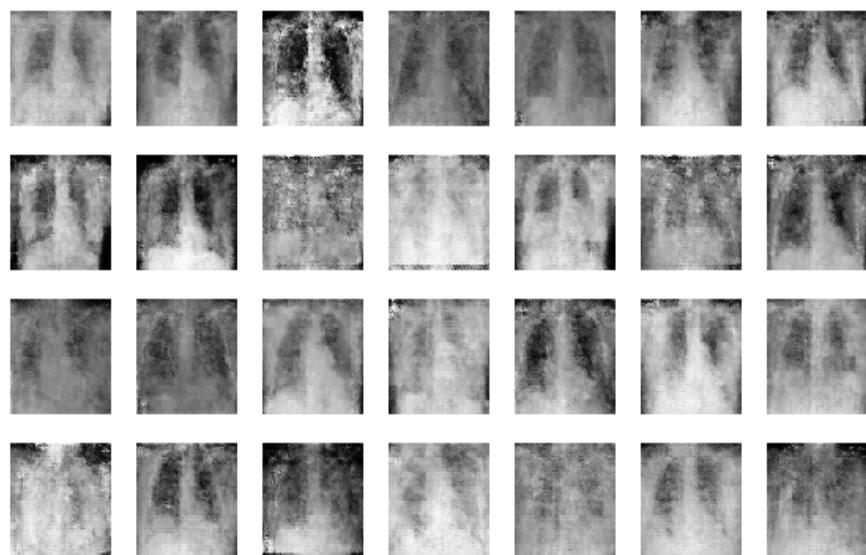


Figure 4.25: 800 Εποχές (64,64,1) - Μικρότερο Μοντέλο

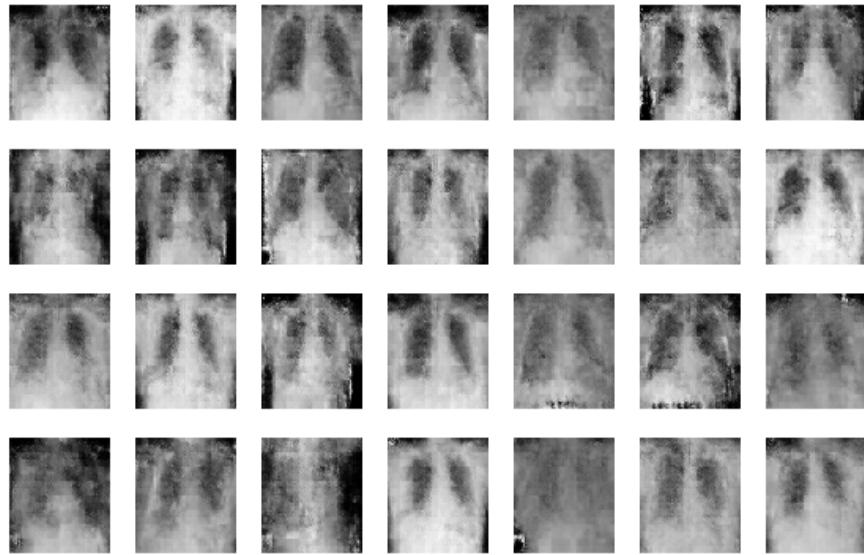


Figure 4.26: 1000 Εποχές (64,64,1) - Μικρότερο Μοντέλο

Παρατηρήθηκε πως το διάγραμμα σφάλματος του Παραγωγικού Δικτύου είχε μια ανοδική πορεία, όπως επίσης το οπτικό αποτέλεσμα δεν ήταν τόσο ικανοποιητικό όσο του μοντέλου που επιλέξαμε. Αυτά αποτελούν σημάδια της Αποτυχίας Σύγχλισης, η οποία ήταν εμφανής, όχι τόσο όμως όσο στα προηγούμενα μοντέλα.

4.1.6 Μοντέλο με μεγαλύτερο αριθμό στρωμάτων

Στο πείραμα αυτό αυξήθηκαν κατά ένα στρώμα το Παραγωγικό και Διαχωριστικό Δίκτυο, δηλαδή τα στρώματα που αφορούν την αύξηση των διαστάσεων (Transpose Convolutional) στο Παραγωγικό Δίκτυο και τα στρώματα που αφορούν στη μείωση των διαστάσεων (Convolutional) στο Διαχωριστικό Δίκτυο. Πιο συγκεκριμένα οι παράμετροι στο Παραγωγικό Δίκτυο άλλαξαν ως εξής :

- Total params: 1,119,841 → 3,003,233
- Trainable params: 1,111,201 → 2,998,177
- Non-trainable params: 8,640 → 5,056

Για το Διαχωριστικό Δίκτυο η αντίστοιχη αλλαγή ήταν η εξής:

- Total params: 26,913,505 → 40,555,745

- Trainable params: 26,912,609 → 40,553,825
- Non-trainable params: 896 → 1,920

Η αύξηση των διαστάσεων έγινε με πιο ομαλό τρόπο, κάτι το οποίο θα ήταν επιθυμητό εάν ήταν διαθέσιμο ένα μεγάλο σετ δεδομένων ικανό να εκπαιδεύσει τις περισσότερες παραμέτρους. Τα αποτελέσματα παρατίθενται παρακάτω 4.1.6 4.1.6 4.1.6 4.1.6 4.1.6.

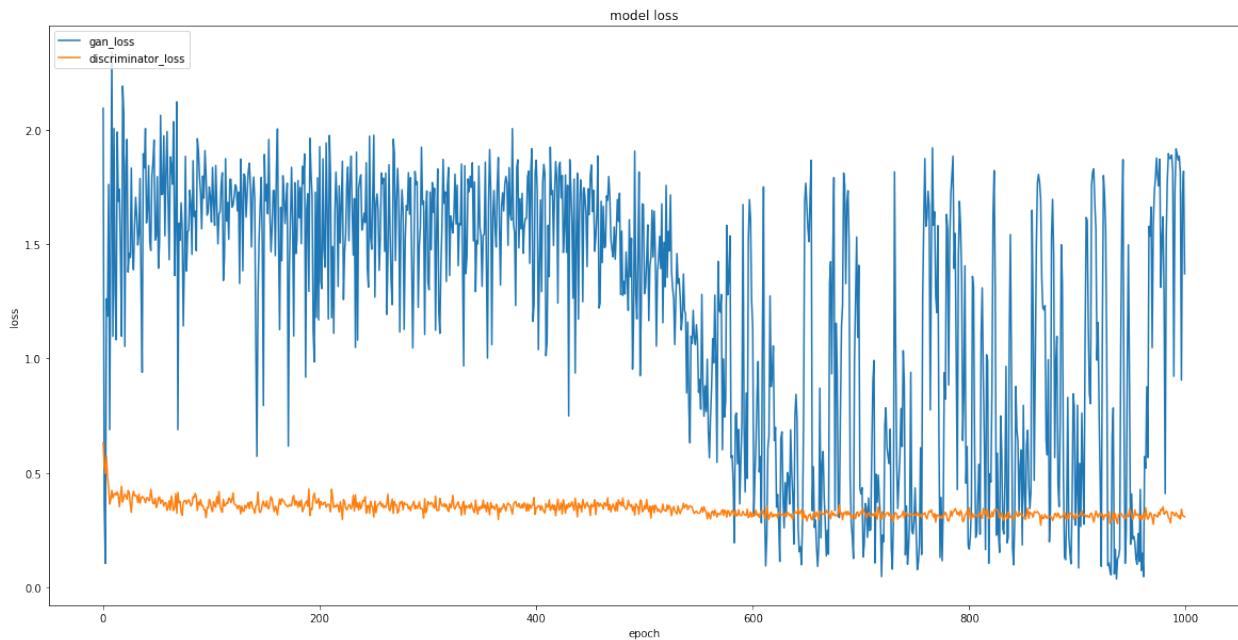


Figure 4.27: Διάγραμμα σφάλματος για (64,64,1) εικόνες - Μεγαλύτερο Μοντέλο

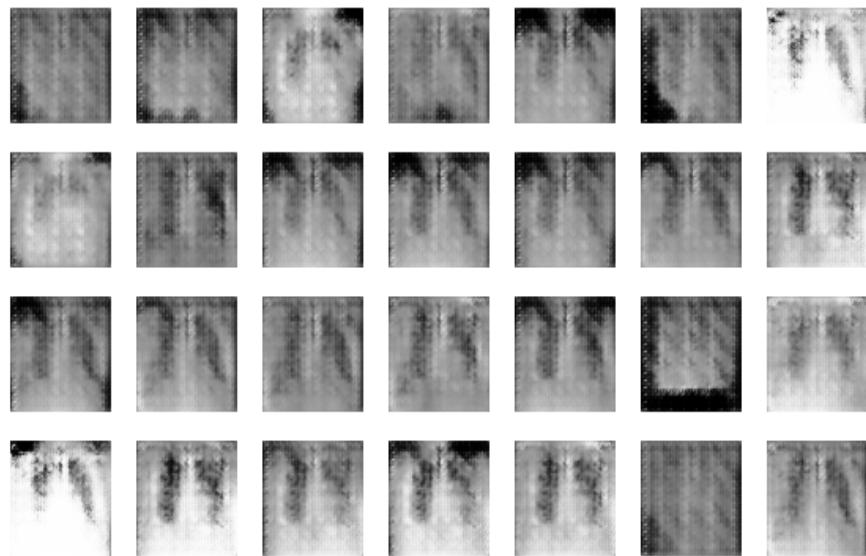


Figure 4.28: 200 Εποχές (64,64,1) - Μεγαλύτερο Μοντέλο

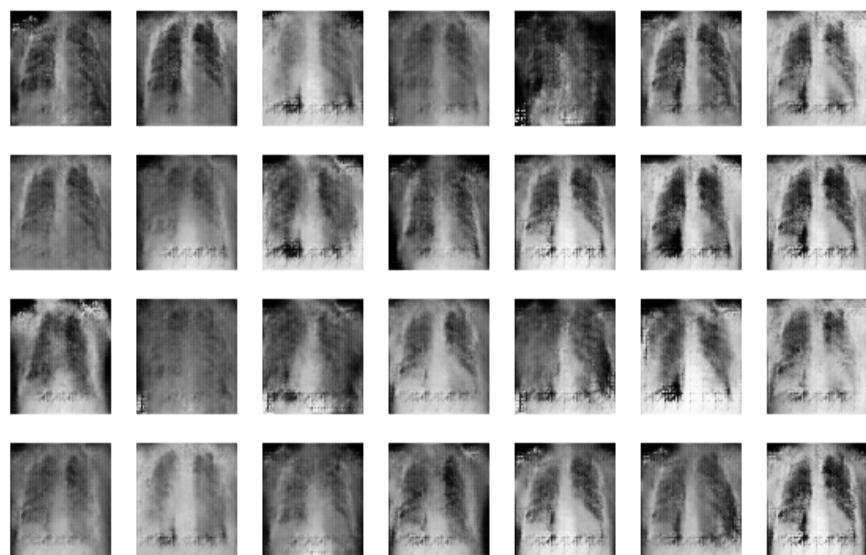


Figure 4.29: 500 Εποχές (64,64,1) - Μεγαλύτερο Μοντέλο

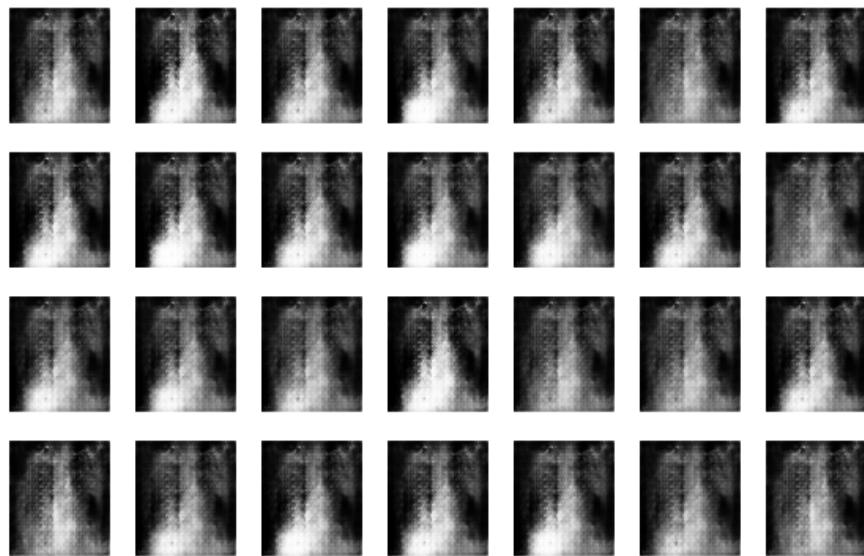


Figure 4.30: 800 Εποχές (64,64,1) - Μεγαλύτερο Μοντέλο

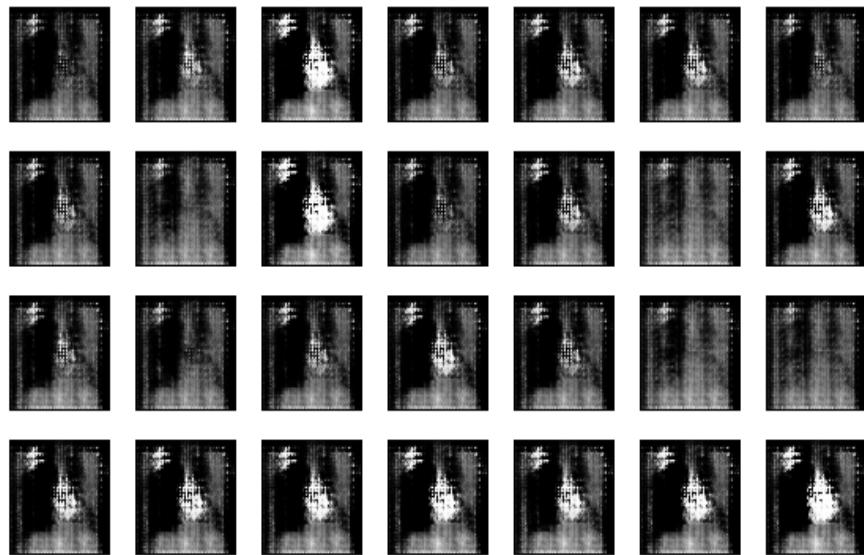


Figure 4.31: 1000 Εποχές (64,64,1) - Μεγαλύτερο Μοντέλο

Το πρόβλημα της Βλάβης Μοντέλου ήταν εμφανές και συσχετίζεται με την χρήση περισσότερων μεταβλητών στη δομή του μοντέλου, κάτι που προέρχεται από την

αύξηση των στρωμάτων.

4.2 Μεταφορά Γνώσης και Εφαρμογή Αλγόριθμου Βελτιστοποίησης Συμήνους Σωματιδίων

Σε αυτή την ενότητα θα παρουσιαστούν τα αποτελέσματα της εφαρμογής της Μεταφοράς Γνώσης και της Επανεκπαίδευσης μέσω του αλγόριθμου Βελτιστοποίησης Συμήνους Σωματιδίων, στα μοντέλα που έχουν επιλεχθεί. Έστερα θα χρησιμοποιηθούν τα αποτελέσματα της ΒΣΣ για να ορισθούν οι βέλτιστες υπερπαράμετροι σε κάθε περίπτωση. Αξιοποιώντας αυτές τις υπερπαραμέτρους, χτίζεται το κάθε μοντέλο και έπειτα εφαρμόζεται επανεκπαίδευση. Ο λόγος για τον οποίο ακολουθείται αυτή τη διαδικασία είναι για να εξακριβωθεί αν η συγκεκριμένη αρχιτεκτονική είναι ικανή να οδηγήσει σε καλά αποτελέσματα, καθώς στην πρώτη εκπαίδευση τα στοχαστικά στοιχεία που προσδίδουν τα Dropout στρώματα - όπως επίσης και ο τυχαίος διαχωρισμός του σετ σε τμήματα εκπαίδευσης και επικύρωσης - μπορεί να οδήγησαν σε πετυχημένη εκπαίδευση που να αποτελεί εξαίρεση. Αν επιτευχθούν παρόμοια επίπεδα εκπαίδευσης, τότε διατηρείται η αρχιτεκτονική και ελέγχεται το διάγραμμα απόδοσης για να αποφασιστεί ο κατάλληλος αριθμός εποχών που δεν οδηγούν σε Υπερπροσαρμογή (Overfit). Η δομή που θα τηρηθεί θα είναι η εξής :

- VGG19 εκπαίδευση
 - VGG19 με Πλήρη Συνδεδεμένα Στρώματα και μικρό dataset (VGG Simple)
 - VGG19 με Πλήρη Συνδεδεμένα Στρώματα και μεγάλο dataset (VGG Enhanced)
 - VGG19 με LSTM Layer και μικρό dataset (VGG LSTM Simple)
 - VGG19 με LSTM Layer και μεγάλο dataset (VGG LSTM Enhanced)
- ResNet152V2 εκπαίδευση
 - ResNet152V2 με Πλήρη Συνδεδεμένα Στρώματα και μικρό dataset (ResNet Simple)
 - ResNet152V2 με Πλήρη Συνδεδεμένα Στρώματα και μεγάλο dataset (ResNet Enhanced)

- ResNet152V2 με LSTM Layer και μικρό dataset (**ResNet LSTM Simple**)
- ResNet152V2 με LSTM Layer και μεγάλο dataset (**ResNet LSTM Enhanced**)
- CheXNet/DenseNet121 εκπαίδευση
 - CheXNet/DenseNet121 με Πλήρη Συνδεδεμένα Στρώματα και μικρό dataset (**CheXNet Simple**)
 - CheXNet/DenseNet121 με Πλήρη Συνδεδεμένα Στρώματα και μεγάλο dataset (**CheXNet Enhanced**)
 - CheXNet/DenseNet121 με LSTM Layer και μικρό dataset (**CheXNet LSTM Simple**)
 - CheXNet/DenseNet121 με LSTM Layer και μεγάλο dataset (**CheXNet LSTM Enhanced**)

Οι παράμετροι που αναγράφονται στους παρακάτω πίνακες είναι οι εξής:

- FL: Frozen Layers, πόσα στρώματα από το τέλος ”παγώνουν”, έτσι ώστε να μην μπορούν να εκπαίδευτούν.
- Fl(1) GAP2D(0): Αμέσως μετά το μοντέλο που επιλέχθηκε χρησιμοποιείται Flatten Layer (αν η τιμή είναι 0) ή GlobalAveragePooling2D Layer (αν η τιμή είναι 1)
- ExtrDL: Extra Dense Layers , πόσα Πλήρη Συνδεδεμένα Στρώματα επιπλέον χρησιμοποιούνται πριν το τελικό classification.
- Dr1 , Dr2: Dropout1,Dropout2 : η τιμή στα Dropout Layers που χρησιμοποιείται μετά από τα 2 τελευταία Πλήρη Συνδεδεμένα Στρώματα.
- LU: LSTM units

Σε όλα τα διαγράμματα απεικονίζεται με **πορτοκαλί** χρώμα η **εκπαίδευση** και με **μπλε** χρώμα η **επικύρωση**.

4.2.1 VGG19

”Φορτώθηκε” το VGG19 μοντέλο, προεκπαίδευμένο με το ImageNet Dataset και εφαρμόστηκε αλγόριθμος ΒΣΣ, ο οποίος αναζητά υπερπαραμέτρους για μεγιστοποίηση της ακρίβειας. Για τη ΒΣΣ χρησιμοποιήθηκαν 20 εποχές στην εκπαίδευση κάθε πιθανού μοντέλου, όπως επίσης και διακοπή εκπαίδευσης (EarlyStopping) με υπομονή 4 εποχές, δηλαδή αν για παραπάνω από 4 εποχές δεν υπήρξε αύξηση της ακρίβειας στο σύνολο επικύρωσης, τότε διακόπτεται η εκπαίδευση και συνεχίζει ο αλγόριθμος την αναζήτηση με άλλες υπερπαραμέτρους. Για κάθε μοντέλο παρουσιάζονται οι υπερπαράμετροι που σχημάτισαν το μοντέλο με την μεγαλύτερη ακρίβεια στο σύνολο επικύρωσης (μοντέλο 1) όπως επίσης και οι υπερπαράμετροι που οδήγησαν στο μοντέλο με το μικρότερο σφάλμα στο σύνολο επικύρωσης (μοντέλο 2)

VGG Simple

Παρατηρήθηκε ότι η μέγιστη ακρίβεια που επιτεύχθηκε είναι στο μοντέλο Rank 1 και σημείωσε **0.907** και το μικρότερο σφάλμα είναι στο Rank 2 και σημείωσε **0.257**. Διακρίνεται επίσης η μεγάλη ομοιότητα στις υπερπαραμέτρους μεταξύ των δύο μοντέλων.

vgg_simple							
Rank	FL	Fl(1) GAP2D(0)	ExrDL	Dr1	Dr2	Accuracy	Loss
1	7	0	2	0.441	0.203	0.907	0.261
2	7	0	2	0.461	0.239	0.901	0.257

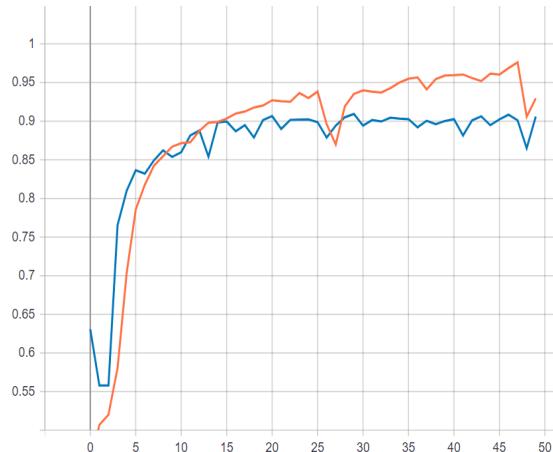


Figure 4.32 vgg_simple 1 Accuracy

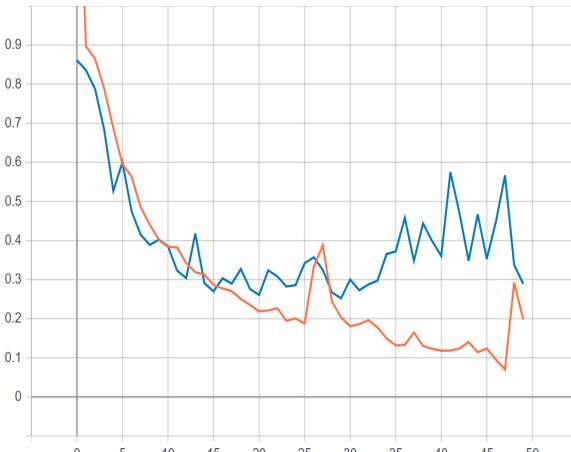


Figure 4.33 vgg_simple 1 Loss

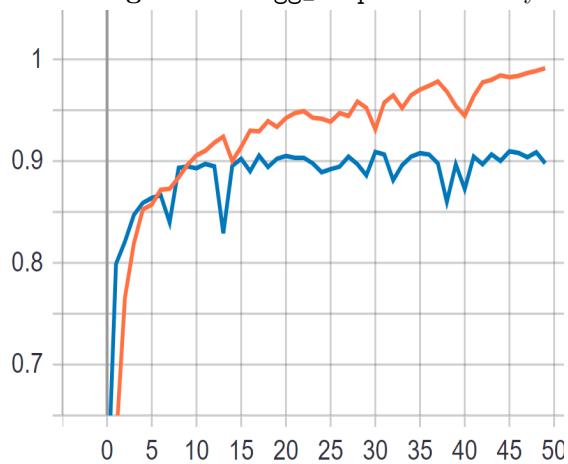


Figure 4.34 vgg_simple 2 Accuracy

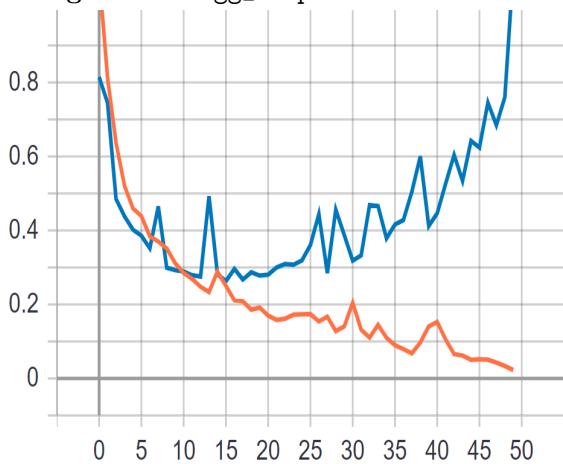


Figure 4.35 vgg_simple 2 Loss

Σημειώνεται πως περίπου στις 20 εποχές το σφάλμα επικύρωσης άρχισε να αυξάνεται, το οποίο είναι σημάδι υπερπροσαρμογής, συνεπώς σταματά η εκπαίδευση των μοντέλων στις 20 εποχές.

VGG Enhanced

Παρατηρήθηκε ότι η μέγιστη ακρίβεια που επιτεύχθηκε είναι στο μοντέλο Rank 1 και σημείωσε **0.928** και το μικρότερο σφάλμα είναι στο Rank 2 και σημείωσε **0.205**. Διακρίνεται επίσης η μεγάλη ομοιότητα στις υπερπαραμέτρους μεταξύ των δύο μοντέλων. Αξίζει να αναφερθεί πως κατά τη διάρκεια της ΒΣΣ υπήρξαν πολλά περισσότερα μοντέλα που ξεπέρασαν το 90% σε ακρίβεια σε αντίθεση με το μικρό dataset που ήταν ελάχιστες περιπτώσεις. Πιο συγκεκριμένα υπήρξαν 79 περιπτώσεις στο ενισχυμένο dataset και 10 περιπτώσεις στο μικρό dataset. Συνεπώς προκύπτει το συμπέρασμα πως ακόμα και αν δεν επιλεχθούν οι βέλτιστες υπερπαραμέτροι μπορεί να σημειωθεί αξιόλογη ακρίβεια.

vgg_enhanced							
Rank	FL	Fl(1) GAP2D(0)	ExrDL	Dr1	Dr2	Accuracy	Loss
1	7	0	1	0	0.380	0.928	0.209
2	7	0	1	0	0.397	0.924	0.205

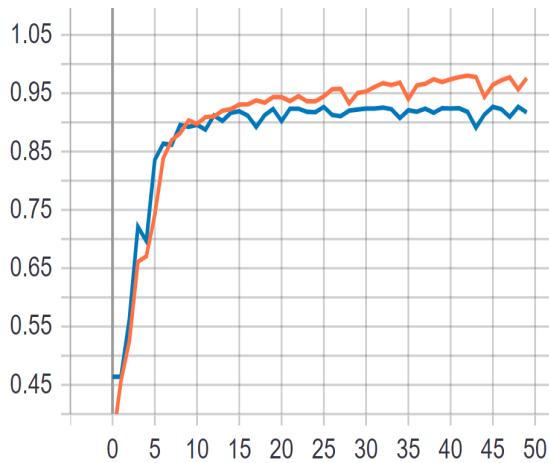


Figure 4.36 vgg_enhanced 1 Accuracy

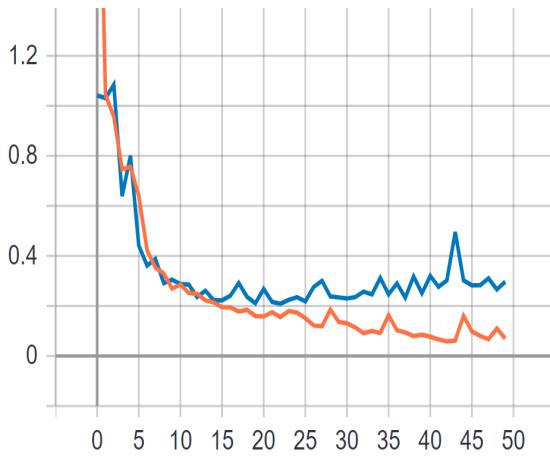


Figure 4.37 vgg_enhanced 1 Loss

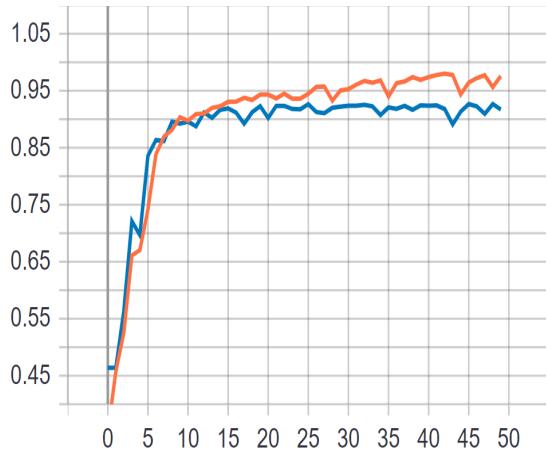


Figure 4.38 vgg_enhanced 2 Accuracy

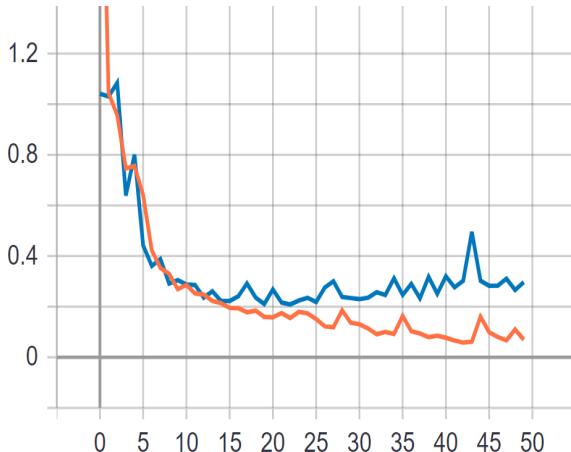


Figure 4.39 vgg_enhanced 2 Loss

Η ίδια παρατήρηση με το vgg_simple εντοπίζεται και σε αυτό το μοντέλο , καθώς στις 20-25 εποχές θα πρέπει να σταματήσει η εκπαίδευση για να μπορέσει το μοντέλο να ανταποκριθεί και σε "ξένα" δεδομένα.

VGG LSTM Simple

Παρατηρήθηκε ότι η μέγιστη ακρίβεια που επιτεύχθηκε είναι στο μοντέλο Rank 1 και σημείωσε **0.924** και το μικρότερο σφάλμα είναι στο Rank 2 και σημείωσε **0.231**.

vgg_lstm_simple					
Rank	FL	LU	Dr1	Accuracy	Loss
1	7	7	0.197	0.924	0.236
2	6	4	0.123	0.919	0.231

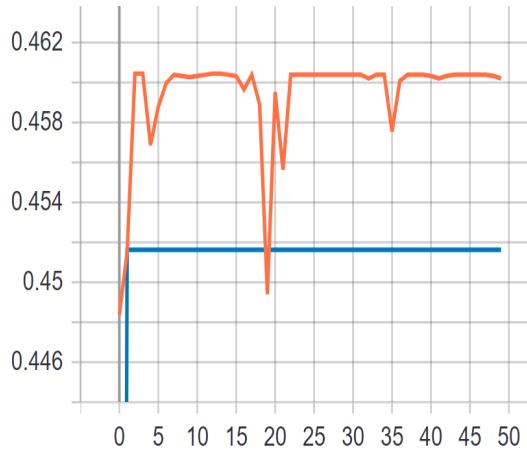


Figure 4.40 vgg_lstm_simple 1
Accuracy

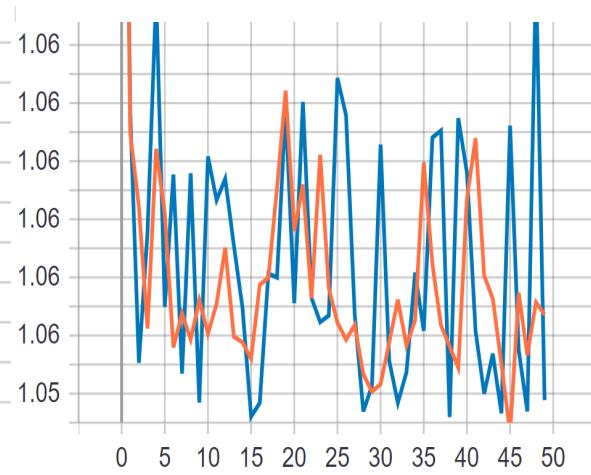


Figure 4.41 vgg_lstm_simple 1 Loss

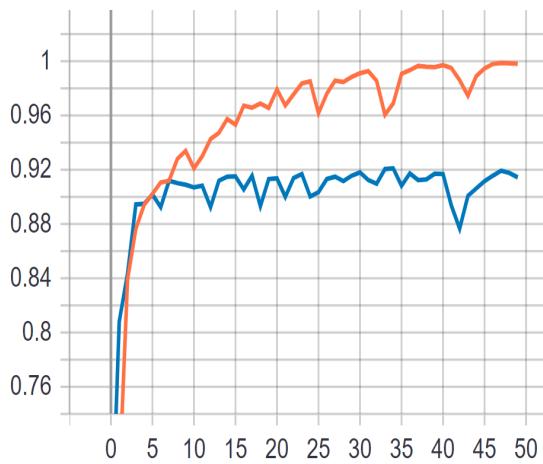


Figure 4.42 vgg_lstm_simple 2
Accuracy

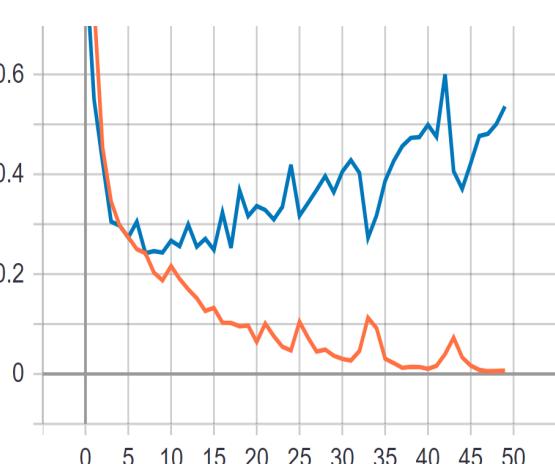


Figure 4.43 vgg_lstm_simple 2 Loss

Εδώ παρατηρήθηκε πως η αρχιτεκτονική του πρώτου μοντέλου δεν μπορεί να εγγυηθεί καλό αποτέλεσμα σε αυτό το dataset εφόσον ελέγχοντας μια διαφορετική κατανομή των δεδομένων προέκυψε κακή εκπαίδευση. Παρόλα αυτά το δεύτερο μοντέλο φαίνεται να επιτυγχάνει την ίδια επίδοση οπότε η αρχιτεκτονική του ανταποκρίνεται σε αυτό το dataset. Τέλος για το δεύτερο μοντέλο εμφανίζεται το φαινόμενο της υπερπροσαρμογής στις 10-15 εποχές.

VGG LSTM Enhanced

Σε αυτή την περίπτωση υπήρξε ταύτιση του μοντέλου με την μεγαλύτερη ακρίβεια και του μοντέλου με το μικρότερο σφάλμα, συνεπώς παρουσιάζεται μόνο μια αρχιτεκτονική. Πιο συγκεκριμένα επιτεύχθηκε **0.926** ακρίβεια και **0.221** σφάλμα.

vgg_lstm_enhanced					
Rank	FL	LU	Dr1	Accuracy	Loss
1	7	6	0.210	0.926	0.221

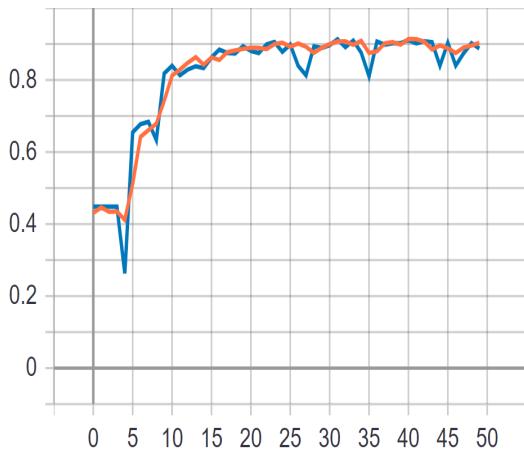


Figure 4.44 vgg_lstm_enhanced 1 Accuracy

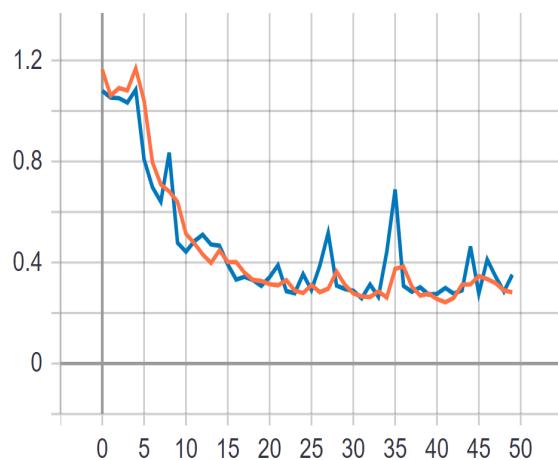


Figure 4.45 vgg_lstm_enhanced 1 Loss

Σημειώνεται ότι περίπου στις 20-25 εποχές σταματάει να προοδεύει το σύστημα, συνεπώς δεν θα χρησιμοποιηθεί μεγαλύτερος αριθμός εποχών

4.2.2 ResNet152V2

”Φορτώθηκε” το ResNet152V2 μοντέλο, προεκπαίδευμένο με το ImageNet Dataset και ακολουθήθηκε η προηγούμενη διαδικασία για 20 εποχές.

ResNet Simple

Παρατηρείται ότι η μέγιστη ακρίβεια που επιτεύχθηκε είναι στο μοντέλο Rank 1 και σημείωσε **0.912** και το μικρότερο σφάλμα είναι στο Rank 2 και σημείωσε **0.543**.

resnet_simple							
Rank	FL	Fl(1) GAP2D(0)	ExrDL	Dr1	Dr2	Accuracy	Loss
1	1	1	0	0.292	0.383	0.912	0.657
2	2	0	0	0.283	0.617	0.895	0.543

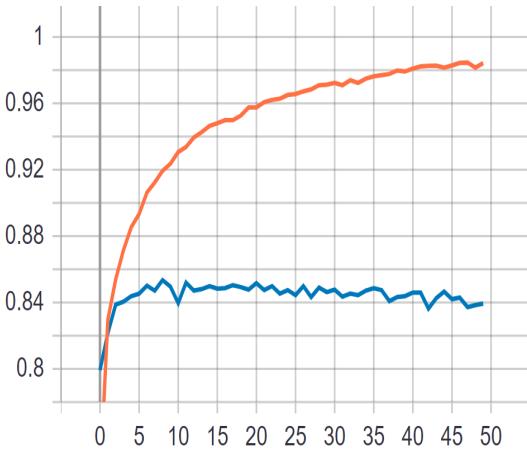


Figure 4.46 resnet_simple 1 Accuracy

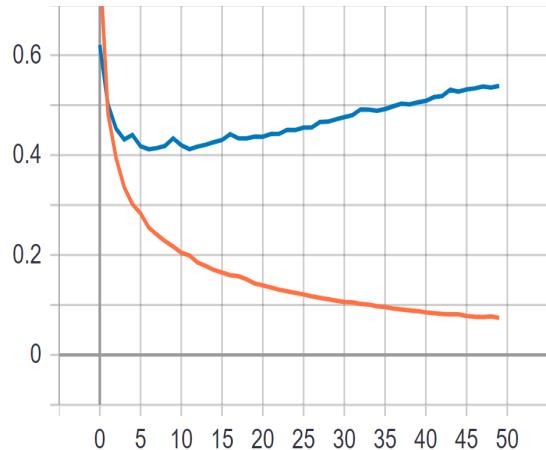


Figure 4.47 resnet_simple 1 Loss

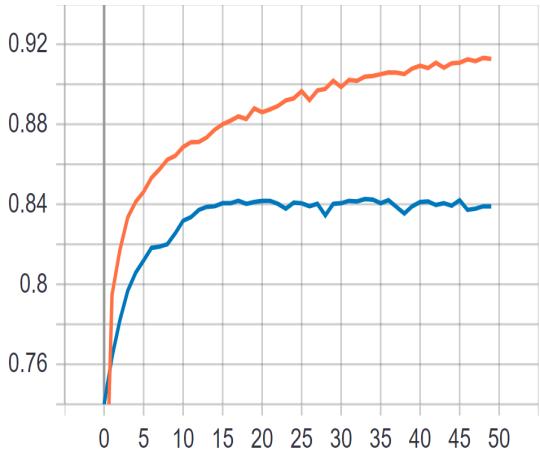


Figure 4.48 resnet_simple 2 Accuracy

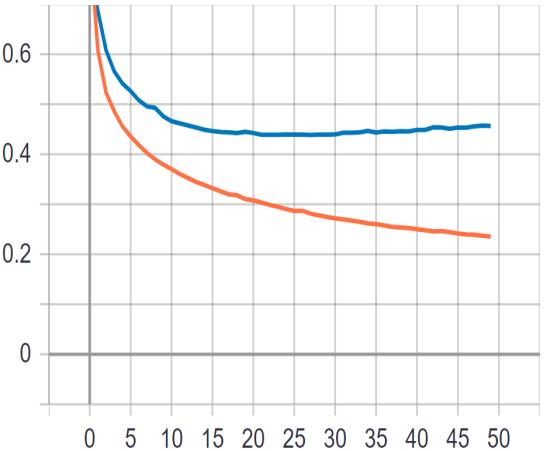


Figure 4.49 resnet_simple 2 Loss

Αυτό που αναδεικνύεται από τα παραπάνω διαγράμματα είναι πως συγχριτικά με τις διάφορες αρχιτεκτονικές του VGG19, σε αυτήν την περίπτωση εμφανίζεται το φαινόμενο της υπερπροσαρμογής πολύ πιο γρήγορα, περί τις 10-15 εποχές.

ResNet Enhanced

Παρατηρείται πως η μέγιστη ακρίβεια που επιτεύχθηκε είναι στο μοντέλο Rank 1 και σημείωσε **0.910** και το μικρότερο σφάλμα είναι στο Rank 2 και σημείωσε **0.375**.

resnet_enhanced							
Rank	FL	Fl(1) GAP2D(0)	ExrDL	Dr1	Dr2	Accuracy	Loss
1	10	0	1	0.470	0.447	0.910	0.603
2	1	1	1	0.533	0.350	0.887	0.375

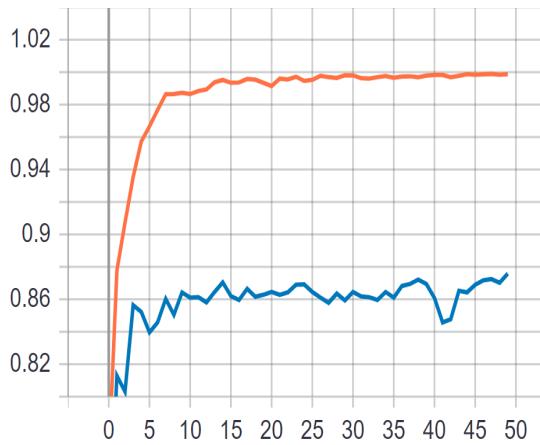


Figure 4.50 resnet_enhanced 1
Accuracy

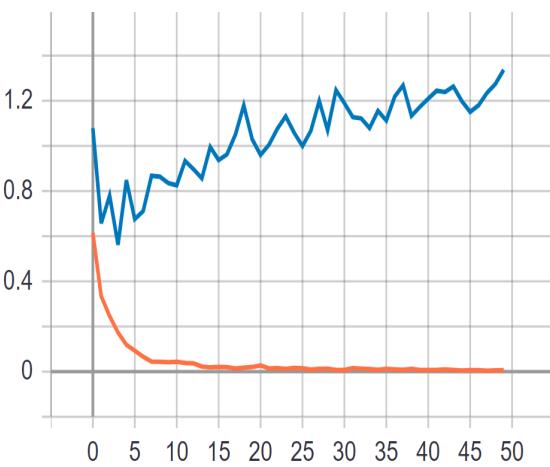


Figure 4.51 resnet_enhanced 1 Loss

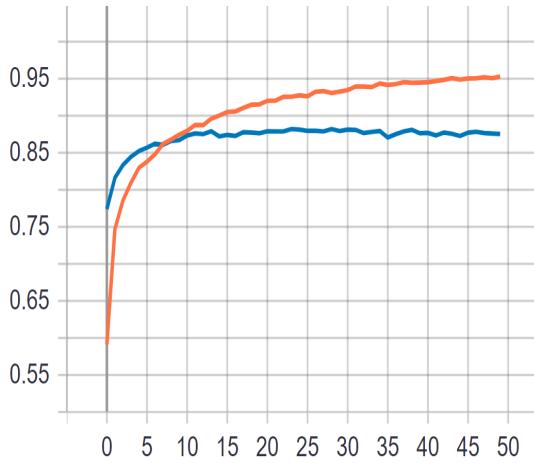


Figure 4.52 resnet_enhanced 2
Accuracy

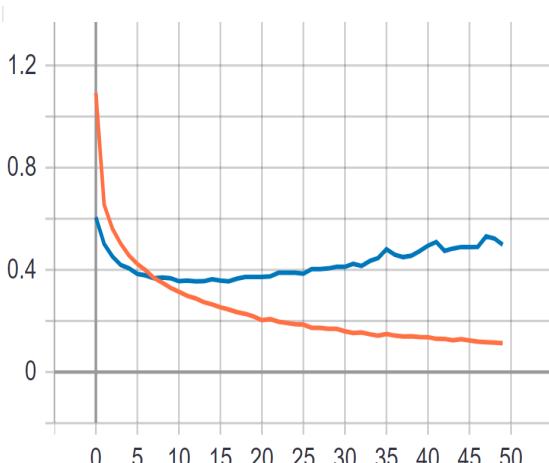


Figure 4.53 resnet_enhanced 2 Loss

Σε αυτήν την περίπτωση διακρίνεται ακόμα πιο έντονο το φαινόμενο της υπερπροσαρμογής, ενώ ο κατάλληλος αριθμός εποχών εκπαίδευσης καθορίζεται στις 5-10.

ResNet LSTM Simple

Παρατηρείται ότι η μέγιστη ακρίβεια που επιτεύχθηκε είναι στο μοντέλο Rank 1 και σημείωσε **0.899** και το μικρότερο σφάλμα είναι στο Rank 2 και σημείωσε **0.671**.

resnet_lstm_simple					
Rank	FL	LU	Dr1	Accuracy	Loss
1	1	10	0.454	0.899	0.700
2	8	3	0.463	0.886	0.671

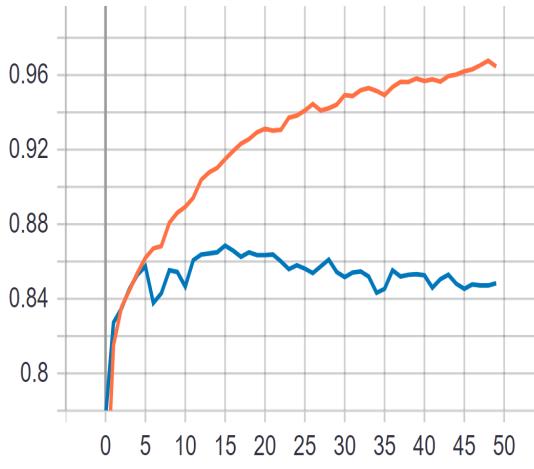


Figure 4.54 resnet_lstm_simple 1
Accuracy

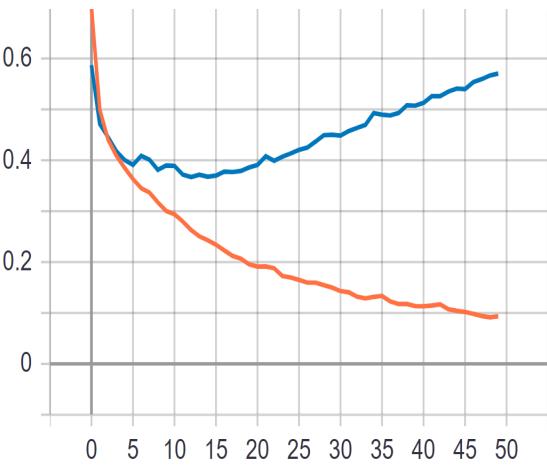


Figure 4.55 resnet_lstm_simple 1 Loss

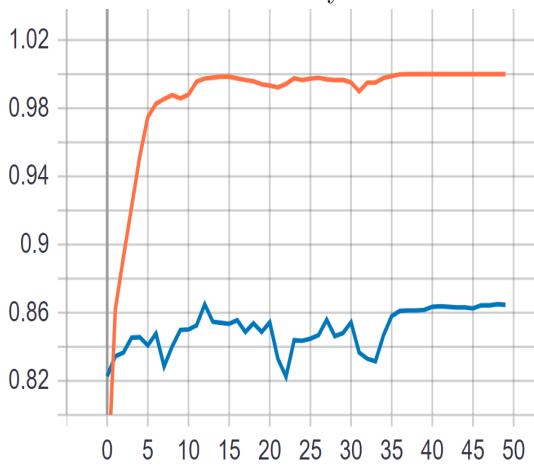


Figure 4.56 resnet_lstm_simple 2
Accuracy

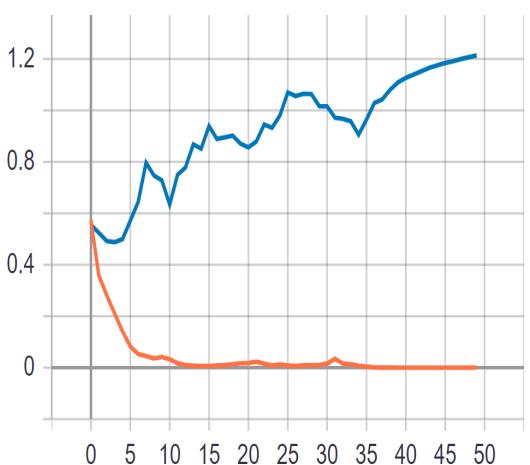


Figure 4.57 resnet_lstm_simple 2 Loss

Στην πρώτη περίπτωση η υπερπροσαρμογή φαίνεται να ξεκινά στην πέμπτη εποχή, ενώ το δεύτερο μοντέλο φαίνεται να μην είναι κατάλληλο για αυτήν την κατανομή του dataset. Η γενικότερη εικόνα που παρουσιάζει το resnet_lstm_simple είναι η αδυναμία του να αντιμετωπίσει επαρκώς το συγκεκριμένο πρόβλημα.

ResNet LSTM Enhanced

Παρατηρείται ότι η μέγιστη ακρίβεια που επιτεύχθηκε είναι στο μοντέλο Rank 1 και σημείωσε **0.914** και το μικρότερο σφάλμα είναι στο Rank 2 και σημείωσε **0.409**.

resnet_lstm_enhanced					
Rank	FL	LU	Dr1	Accuracy	Loss
1	9	3	0.133	0.914	0.562
2	1	1	0.420	0.853	0.409

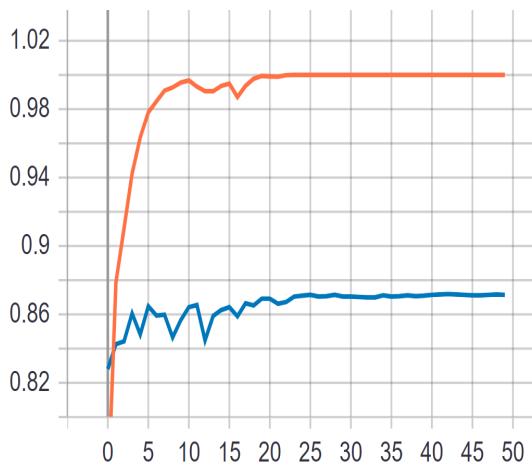


Figure 4.58 resnet_lstm_enhanced 1 Accuracy

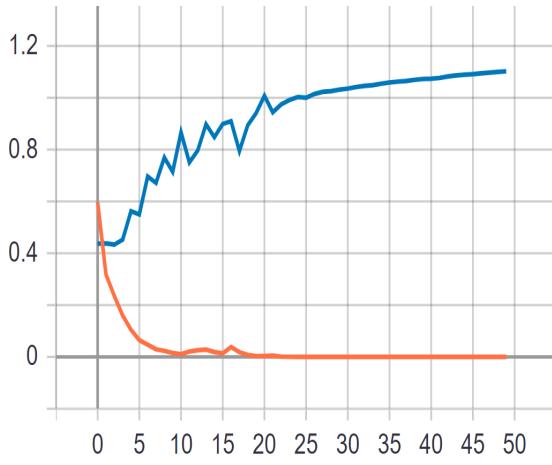


Figure 4.59: resnet_lstm_enhanced 1 Loss

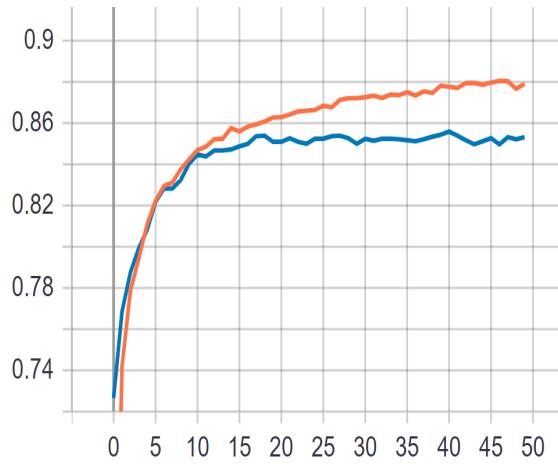


Figure 4.60 resnet_lstm_enhanced 2 Accuracy

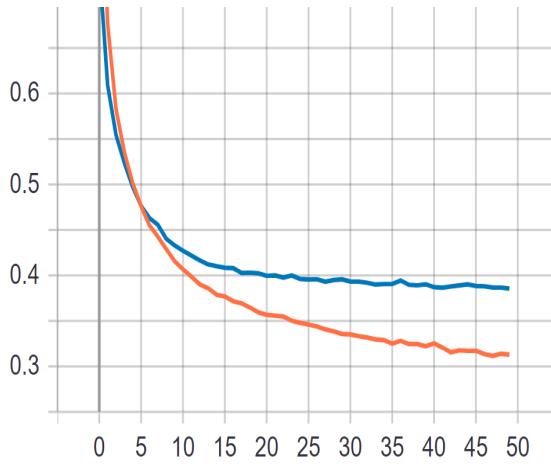


Figure 4.61 resnet_lstm_enhanced 2 Loss

Μια ανάλογη εικόνα με το `resnet_lstm_simple` παρουσιάζει το πρώτο μοντέλο του `resnet_lstm_enhanced`, παρόλα αυτά το δεύτερο μοντέλο παρουσιάζει μια πιο ομαλή εκπαίδευση που φαίνεται να ταιριάζει καλύτερα σε αυτό το σετ δεδομένων. Διαχρίνεται από το διάγραμμα ότι δεν εμφανίζεται το πρόβλημα της υπερπροσαρμογής, ενώ παράλληλα το σφάλμα εγκυρώτητας (validation error) εμφανίζει σταδιακή μείωση.

4.2.3 CheXNet/Dense121

“Φορτώθηκε” το CheXNet μοντέλο, και εφαρμόστηκαν τα βάρη τα οποία προέκυψαν από την εκπαίδευση με το σετ ακτινογραφιών πνεύμονα. Στη συνέχεια, επαναλήφθηκε η διαδικασία που αξιοποιήθηκε στα δύο προηγούμενα μοντέλα, αυτήν τη φορά για 50 εποχές.

CheXNet Simple

Σε αυτή την περίπτωση το μοντέλο με την μεγαλύτερη ακρίβεια ταυτίζεται με το μοντέλο με το μικρότερο σφάλμα, συνεπώς παρουσιάζεται μόνο μια αρχιτεκτονική. Πιο συγκεκριμένα, πέτυχε **0.890** ακρίβεια και **0.415** σφάλμα.

chexnet_simple							
Rank	FL	Fl(1) GAP2D(0)	ExrDL	Dr1	Dr2	Accuracy	Loss
1	120	0	0	0.107	0.270	0.890	0.415

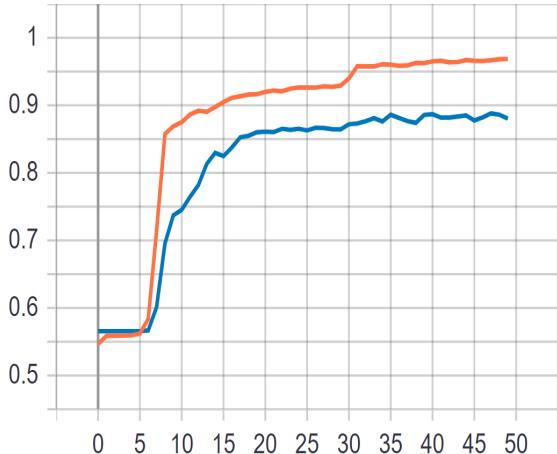


Figure 4.62 chexnet_simple 1 Accuracy

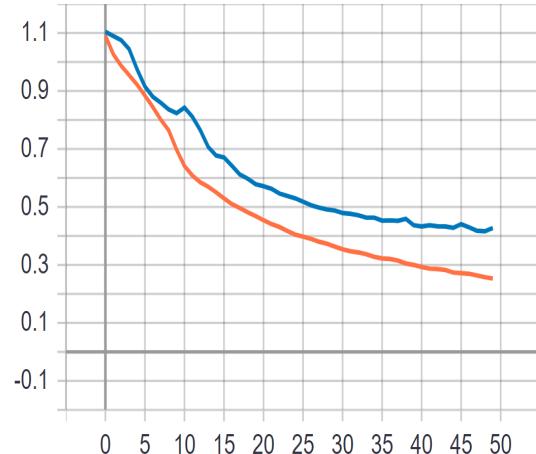


Figure 4.63 chexnet_simple 1 Loss

Σε αυτήν την περίπτωση, το φανόμενο της υπερπροσαρμογής δεν εμφανίστηκε, συνεπώς δεν απαιτήθηκε η δοκιμή του ίδιου μοντέλου με διαφορετική εκπαίδευση.

CheXNet Enhanced

Παρατηρείται ότι η μέγιστη ακρίβεια που επιτεύχθηκε είναι στο μοντέλο Rank 1 και σημειώσε **0.903** και το μικρότερο σφάλμα είναι στο Rank 2 και σημειώσε **0.413**. Αξίζει να σημειωθεί πως η BSS στη συγκεκριμένη περίπτωση κατάφερε να βγάλει το μεγαλύτερο πλήθος μοντέλων με ακρίβεια πάνω από 0.86 με 104 διαφορετικές αρχιτεκτονικές, γεγονός που ενθαρρύνει τη χρήση του μεγάλου dataset.

chexnet_enhanced							
Rank	FL	Fl(1) GAP2D(0)	ExrDL	Dr1	Dr2	Accuracy	Loss
1	120	0	1	0.132	0.422	0.903	0.413

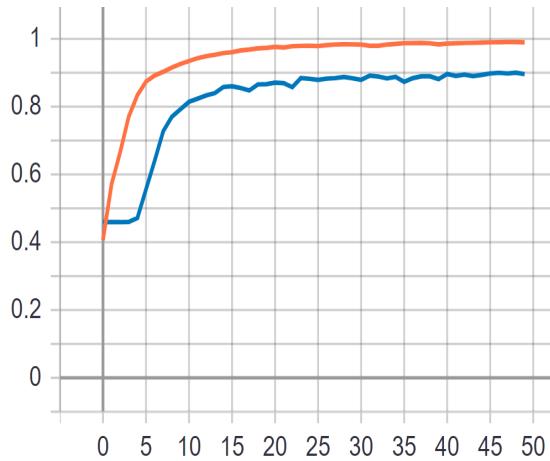


Figure 4.64 chexnet_enhanced 1
Accuracy

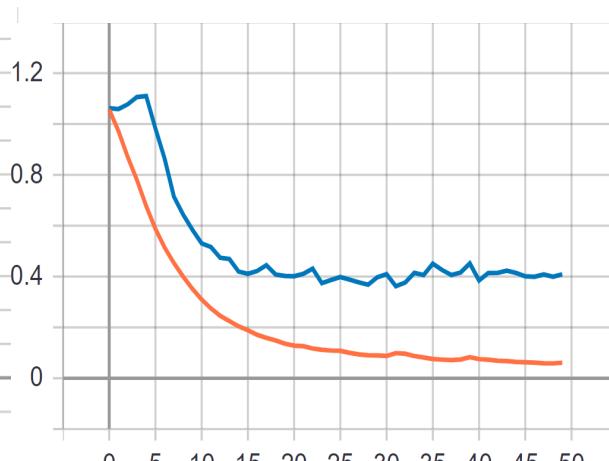


Figure 4.65 chexnet_enhanced 1 Loss

Παρατηρείται σταθεροποίηση του σφάλματος και όχι αύξηση, η οποία θα μπορούσε να θεωρηθεί επικίνδυνη για υπερπροσαρμογή, ενώ παράλληλα σημειώνεται μια σταδιακή αύξηση της ακρίβειας, συνεπώς ο κατάλληλος αριθμός εποχών κυμαίνεται γύρω στις 50.

CheXNet LSTM Simple

Παρατηρείται ότι η μέγιστη ακρίβεια που επιτεύχθηκε είναι στο μοντέλο Rank 1 και σημείωσε **0.882** και το μικρότερο σφάλμα είναι στο Rank 2 και σημείωσε **0.475**.

chexnet_lstm_simple					
Rank	FL	LU	Dr1	Accuracy	Loss
1	120	5	0.6	0.882	0.536
2	120	5	0.303	0.879	0.475

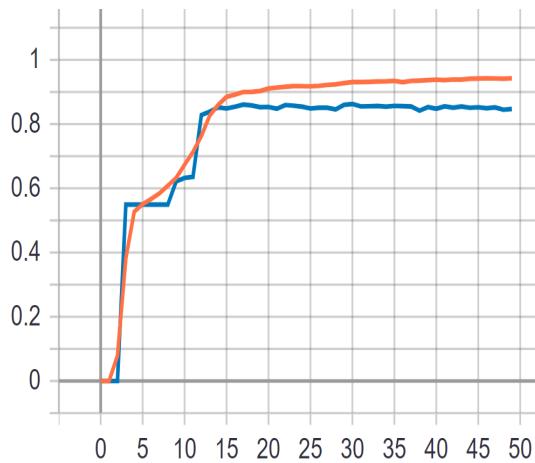


Figure 4.66 chexnet_lstm_simple 1 Accuracy

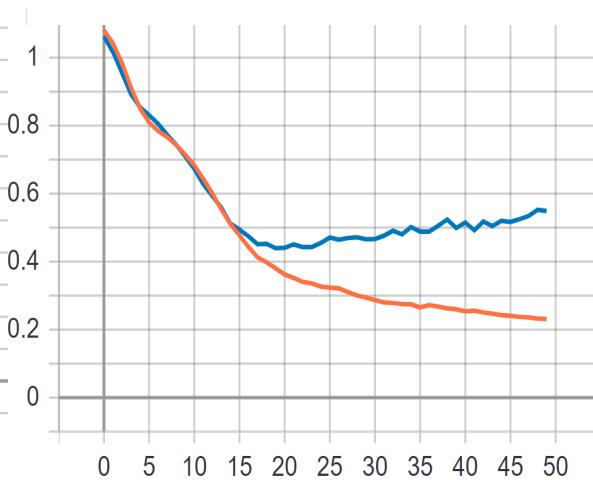


Figure 4.67 chexnet_lstm_simple 1 Loss

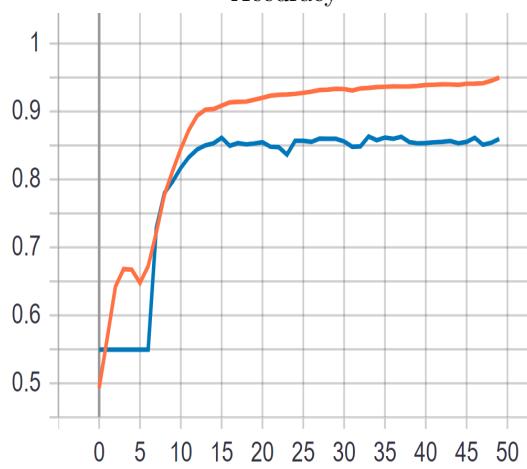


Figure 4.68 chexnet_lstm_simple 2 Accuracy

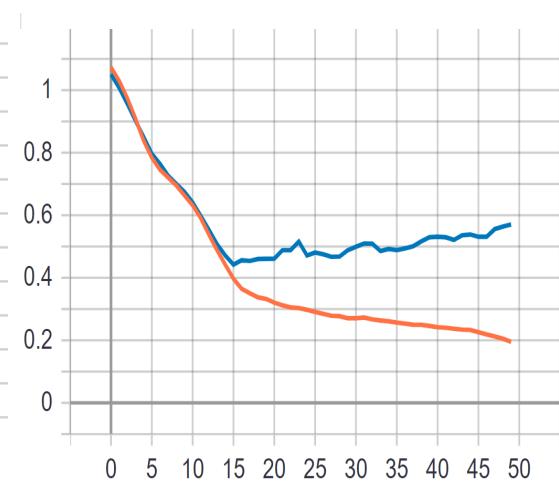


Figure 4.69 chexnet_lstm_simple 2 Loss

Σε αυτές τις αρχιτεκτονικές παρατηρείται το φαινόμενο της υπερπροσαρμογής. Σύμφωνα με τα διαγράμματα, ο καταλληλότερος αριθμός εποχών για εκπαίδευση ορίζεται στα 20.

CheXNet LSTM Enhanced

Σε αυτή την περίπτωση το μοντέλο με την μεγαλύτερη ακρίβεια ταυτίζεται με το μοντέλο με το μικρότερο σφάλμα, συνεπώς παρουσιάζεται μόνο μια αρχιτεκτονική. Πιο συγκεκριμένα πέτυχε **0.906** ακρίβεια και **0.386** σφάλμα.

chexnet_lstm_enhanced					
Rank	FL	LU	Dr1	Accuracy	Loss
1	120	1	0.153	0.906	0.86

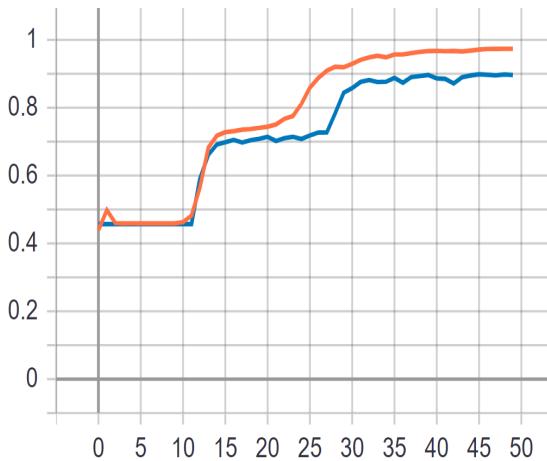


Figure 4.70 chexnet_lstm_enhanced 1
Accuracy

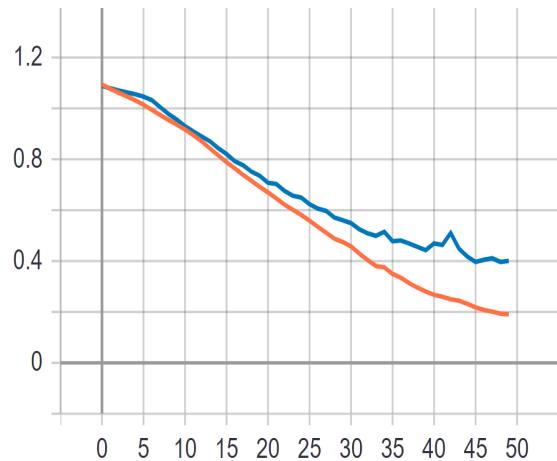


Figure 4.71 chexnet_lstm_enhanced 1
Loss

Διαφαίνεται από τα διαγράμματα ότι οι 50 εποχές αποτελούν έναν αρκετά κατάλληλο αριθμό για εκπαίδευση.

4.3 Τελικά αποτελέσματα

Στις προηγούμενες ενότητες επιχειρήθηκε η δημιουργία μιας συλλογής με τα βέλτιστα μοντέλα. Σε αυτή την ενότητα όμως γίνεται η τελική εκτίμηση μέσω του συνόλου ελέγχου για να αξιολογηθεί η επίδοση τους σε "ξένα" δεδομένα.

Εφαρμόζοντας ΒΣΣ, επιλέχθηκαν ορισμένες αρχιτεκτονικές και προέκυψαν συμπεράσματα για τις εποχές εκπαίδευσης χάθε περίπτωσης. Σε ορισμένες περιπτώσεις, οι εποχές που χρησιμοποιήθηκαν για την αναζήτηση της ΒΣΣ οδηγούσαν σε υπερμοντελοποίηση. Πιο συγκεκριμένα, οι 20 εποχές για το μοντέλο VGG19 χρίθηκαν ως ένας αριθμός που δεν θα δημιουργήσει πρόβλημα. Αντίθετα, για το μοντέλο ResNet15V2 οι 20 εποχές χρίθηκαν ως ένας αριθμός που μπορεί να οδηγήσει σε υπερμοντελοποίηση και η πρόταση διαφορφώθηκε στις 10 εποχές. Για το CheXnet/DenseNet μοντέλο, οι 50 εποχές δημιουργησαν πρόβλημα σε 2 περιπτώσεις με την τελική πρόταση να είναι οι 20 εποχές. Στα μοντέλα στα οποία δημιουργήθηκε πρόβλημα πραγματοποιήθηκε η εκπαίδευση με τον αντίστοιχο αριθμό εποχών και καταγράφηκε η ακρίβεια την οποία πετυχαίνουν στο σύνολο επικύρωσης - αν η ακρίβεια είναι μεγαλύτερη του 85% τότε το μοντέλο συγκαταλέγεται στην συλλογή.

Η τελική συλλογή αποτελείται από:

- 20 epochs: vgg_simple 1, vgg_simple 2, vgg_enhanced 1, vgg_enhanced 2, vgg_lstm_simple 1, vgg_lstm_simple 2, vgg_lstm_enhanced 1
- 20 epochs: resnet_simple 1, resnet_simple 2, resnet_enhanced 1, resnet_enhanced 2, resnet_lstm_simple 1, resnet_lstm_simple 2, resnet_lstm_enhanced 1, resnet_lstm_enhanced 2
- 10 epochs: resnet_simple 2, resnet_enhanced 2, resnet_lstm_simple 1, resnet_lstm_simple 2, resnet_lstm_enhanced 2
- 50 epochs: chexnet_simple1, chexnet_enhanced1, chexnet_lstm_simple1, chexnet_lstm_simple2, chexnet_lstm_enhanced
- 20 epochs: chexnet_lstm_simple1, chexnet_lstm_simple2

Στον πίνακα φαίνονται τα τελικά αποτελέσματα:

	VGG19	ResNet152 V2 (20 Epochs)	ResNet152 V2 (10 Epochs)	CheXNet/ DenseNet (50 Epochs)	CheXNet/ DenseNet (20 Epochs)
simple1	0.90	0.89	-	0.87	-
simple2	0.90	0.88	0.85	-	-
enhanced1	0.90	0.88	-	0.88	-
enhanced2	0.89	0.84	0.84	-	-
Istm_simple1	0.90	0.88	0.83	0.87	0.84
Istm_simple2	0.90	0.88	0.83	0.87	0.84
Istm_enhanced1	0.90	0.89	-	0.87	-
Istm_enhanced2	-	0.83	0.82	-	-

Table 4.13: Ακρίβεια στο Τεστ Σετ

Από τα αποτελέσματα του πίνακα, επιλέχθηκαν αυτά τα οποία έχουν την κορυφαία ακρίβεια στο VGG19 (6 μοντέλα), στο ResNet152V2 (2 μοντέλα) και στο CheXNet/DenseNet (1 μοντέλο) και συγχροτήθηκε μια πιο αναλυτική αναφορά με επιπλέον μετρικές, που θα ώστε να διαμορφωθεί μια καλύτερη εικόνα για το πώς κάθε μοντέλο αντιμετώπισε το σύνολο ελέγχου.

	precision	recall	f1-score	support
covid19	0.79	0.57	0.67	80
normal	0.91	0.92	0.92	905
pneumonia	0.90	0.91	0.90	800
accuracy			0.90	1785
macro avg	0.87	0.80	0.83	1785
weighted avg	0.90	0.90	0.90	1785

Table 4.14: vgg_simple 1

	precision	recall	f1-score	support
covid19	0.74	0.69	0.71	80
normal	0.89	0.95	0.92	905
pneumonia	0.94	0.87	0.90	800
accuracy			0.90	1785
macro avg	0.86	0.84	0.85	1785
weighted avg	0.90	0.90	0.90	1785

Table 4.15: vgg_simple 2

	precision	recall	f1-score	support
covid19	0.83	0.66	0.74	80
normal	0.92	0.90	0.91	905
pneumonia	0.88	0.92	0.90	800
accuracy			0.90	1785
macro avg	0.88	0.83	0.85	1785
weighted avg	0.90	0.90	0.90	1785

Table 4.16: vgg_enhanced 1

	precision	recall	f1-score	support
covid19	0.81	0.71	0.76	80
normal	0.90	0.94	0.92	905
pneumonia	0.92	0.88	0.90	800
accuracy			0.90	1785
macro avg	0.88	0.84	0.86	1785
weighted avg	0.90	0.90	0.90	1785

Table 4.17: vgg_lstm_simple 1

	precision	recall	f1-score	support
covid19	0.81	0.62	0.70	80
normal	0.90	0.94	0.92	905
pneumonia	0.90	0.88	0.89	800
accuracy			0.90	1785
macro avg	0.87	0.81	0.84	1785
weighted avg	0.90	0.90	0.90	1785

Table 4.18: vgg_lstm_simple 2

	precision	recall	f1-score	support
covid19	0.70	0.68	0.69	80
normal	0.92	0.92	0.92	905
pneumonia	0.90	0.90	0.90	800
accuracy			0.90	1785
macro avg	0.84	0.83	0.84	1785
weighted avg	0.90	0.90	0.90	1785

Table 4.19: vgg_lstm_enhanced 1

	precision	recall	f1-score	support
covid19	0.89	0.68	0.77	80
normal	0.88	0.93	0.91	905
pneumonia	0.90	0.88	0.89	800
accuracy			0.89	1785
macro avg	0.89	0.83	0.85	1785
weighted avg	0.89	0.89	0.89	1785

Table 4.20: resnet_simple 1

	precision	recall	f1-score	support
covid19	0.80	0.69	0.74	80
normal	0.90	0.90	0.90	905
pneumonia	0.88	0.89	0.89	800
accuracy			0.89	1785
macro avg	0.86	0.83	0.84	1785
weighted avg	0.89	0.89	0.89	1785

Table 4.21: resnet_lstm_enhanced 1

	precision	recall	f1-score	support
covid19	0.77	0.69	0.73	80
normal	0.89	0.90	0.89	905
pneumonia	0.87	0.88	0.87	800
accuracy			0.88	1785
macro avg	0.85	0.82	0.83	1785
weighted avg	0.88	0.88	0.88	1785

Table 4.22: chexnet_enhanced 1

Η συγκεκριμένη εργασία επικεντρώθηκε ιδιαίτερα στην εύρεση εικόνων ασθενών με COVID-19 ενώ παράλληλα δόθηκε μεγάλη έμφαση στην παραγωγή πλασματικών, αλλά όσο το δυνατό πιο ρεαλιστικών εικόνων ασθενών με COVID-19. Τέλος συνδυάστηκαν τα παραπάνω δύο με σκοπό να επιτευχθεί η καλύτερη ανίχνευση του. Επίσης είναι κατανοητό πως ένα αρνητικό δείγμα που δεν ανιχνεύθηκε είναι πιο επικίνδυνο από ένα αρνητικό δείγμα το οποίο κρίθηκε ως θετικό. Από τα παραπάνω προκύπτει το συμπέρασμα πως η μετρική

Recall θα έχει μεγαλύτερη βαρύτητα στην τελική επίλογή του μοντέλου μας. Υστερα από όλη αυτή τη μελέτη το μοντέλο με την καλύτερη Recall τιμή για την κλάση covid19 είναι το vgg_lstm_simple 1 με **0.71** και για την κλάση pneumonia είναι το vgg_enhanced 1 με **0.92**.

Το μοντέλο VGG19 με το μικρότερο βάθος αλλά μεγαλύτερο αριθμό εκπαιδεύσιμων παραμέτρων συγχριτικά με τα υπόλοιπα μοντέλα, υπερίσχυσε και κατάφερε να ξεχωρίσει σε όλες τις περιπτώσεις και αρχιτεκτονικές. Σαν βέλτιστα τελικά μοντέλα παρουσιάστηκαν οι προτάσεις μας στην παραπόνω παράγραφο, αλλά είναι αρκετά σημαντικό να αναφερθεί πως το ενισχυμένο σε δεδομένων κατάφερε με πολλούς διαφορετικούς συνδυασμούς υπερπαραμέτρων να επιτύχει αρκετά ικανοποιητικά αποτελέσματα κατά την εκπαίδευση του. Αυτό μας οδηγεί στο συμπέρασμα ότι ακόμα και αν η ρύθμιση των υπερπαραμέτρων δεν είναι η βέλτιστη, το ενισχυμένο σε δέσμη περισσότερες πιθανότητες να εκπαιδεύσει ένα μοντέλο που να επιτυγχάνει υψηλή ακρίβεια. Συνεπώς σε περιπτώσεις που η ρύθμιση υπερπαραμέτρων δεν είναι εφικτή λόγω υπολογιστικών πόρων ή διαθέσιμου χρόνου, σαν βέλτιστη επιλογή προτείνεται το vgg_enhanced

4.4 Σημασία Ποιότητας και Ποσότητας στα Δεδομένα

Όλη η διαδικασία εύρεσης κατάλληλης αρχιτεκτονικής για κάθε προεκπαίδευμένο μοντέλο στηρίζεται στα χαρακτηριστικά των δεδομένων, όπως επίσης και στους πόρους που μπορούν να διατεθούν τόσο για την εκπαίδευση όσο και για την αποθήκευση.

Το βασικό χαρακτηριστικό του ενισχυμένου σε δεδομένων είναι η ανάλυση των εικόνων, η οποία είναι (64,64,1). Στο αρχικό σε δεδομένων υπήρχαν και εικόνες με (1024,1024,3) ανάλυση, επομένως πρόκειται για μια μεγάλη υποβάθμιση, κατά την οποία διακινδυνεύεται η ποιότητα και μαζί με αυτή αρκετές πληροφορίες για τις εικόνες όπως και χαρακτηριστικά τους.

Η ανάλυση αυτή αποφασίστηκε για δύο λόγους. Πρώτος λόγος, ήταν ο αποθηκευτικός χώρος - οι διαθέσιμοι πόροι δεν επέτρεπαν να "φορτωθεί" το συγκεκριμένο σε δεδομένων και να πραγματοποιηθεί μια εκπαίδευση μοντέλου για μεγάλες αναλύσεις. Δοκιμάστηκε μέχρι και (180,180,3) και απέτυχε καθώς η μνήμη RAM δεν επαρκούσε. Ο δεύτερος λόγος ήταν συνέπεια του πρώτου λόγου, καθώς όπως έχει αναφερθεί στο κεφάλαιο 4.1 για εικόνες χαμηλών αναλύσεων- που επέτρεπαν την αποθήκευσή τους- το ΠΑΔ που κατάφερε να ξεχωρίσει αντιστοιχούσε σε παραγωγή εικόνων (64,64,1).

Παρόλα αυτά, διερευνήθηκε αν έχει μεγαλύτερη σημασία σε ένα σε δεδομένων, όπως αυτό που αξιοποιήθηκε, η ενίσχυση του σε με **ποσότητα** αλλά με χειρότερη ανάλυση ή η ενίσχυση του σε με **ποιότητα**, κάτι που θα είχε όμως αντίκτυπο στο μέγεθος του. Στην τελευταία περίπτωση, εφόσον για καλύτερη ανάλυση δεν γίνεται να εκπαιδευτεί το ΠΑΔ ικανοποιητικά, δεν είναι δυνατή η χρήση "πλασματικών" εικόνων.

Αποφασίστηκε η δοκιμή ενός σε δεδομένων με ανάλυση (128,128,1), όπου η μόνη ενέργεια για να αυξηθεί το πλήθος ήταν να διπλασιαστούν τα δεδομένα, εφαρμόζοντας ακριβώς την ίδια διαδικασία Επαύξησης Δεδομένων που αξιοποιήθηκε και στο ενισχυμένο σε που έχει αναλυθεί. Έπειτα, καθώς τα προεκπαίδευμένα μοντέλα προϋπέθεταν η είσοδος

τους να είναι εικόνες με 3 κανάλια χρωμάτων, επαναλαμβάνεται 3 φορές το αρχικό κανάλι χρώματος οδηγώντας σε ανάλυση (128,128,3).



Figure 4.72: Ανάλυση (128,128,3)



Figure 4.73: Ανάλυση (64,64,3)

Το νέο σετ δεδομένων εικόνων αποτελείται από **1364** εικόνες κλάσης covid19 συγχριτικά με τις **4364** εικόνες κλάσης covid19 του ενισχυμένου σετ

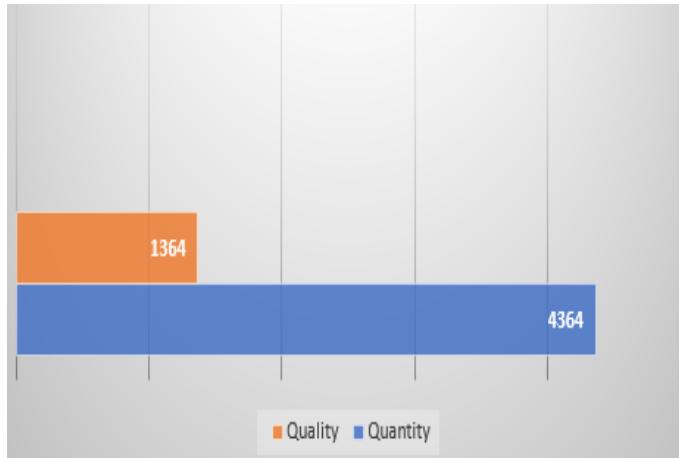


Figure 4.74: Ποσότητα - Ποιότητα

Έχοντας βρει στις παραπάνω παραγράφους τις βέλτιστες υπερπαραμέτρους για κάθε περίπτωση μοντέλου, θα χρησιμοποιηθούν αυτές οι τιμές και θα κατασκευαστούν ακριβώς τα ίδια μοντέλα. Έπειτα, θα εκπαιδευτούν αυτά τα μοντέλα με το νέο σετ δεδομένων που αποτελείται από εικόνες με καλύτερη ανάλυση. Ο λόγος που δεν εφαρμόζεται ο αλγόριθμος ΒΣΣ, είναι πάλι λόγω των πόρων, τόσο σε μνήμη όσο και χρόνο. Παρακάτω παρατίθενται τα αποτελέσματα.

Table 4.23: Ακρίβεια στο Τεστ Σετ

	VGG19	ResNet152V2	CheXNet/ DenseNet
simple1	0.93	0.90	0.89
simple2	0.93	0.89	-
enhanced1	0.93	0.89	0.91
enhanced2	0.92	0.89	-
Istm_simple1	0.5	0.89	0.89
Istm_simple2	0.90	0.90	0.89
Istm_enhanced1	0.89	0.89	0.89
Istm_enhanced2	-	0.85	-

Εντοπίζεται μια σημαντική αύξηση στην ακρίβεια ,χυρίως στο VGG19 μοντέλο, της τάξεως του 3%. Παρουσιάζονται αναλυτικά οι μετρικές που πρόσκυψαν από την εκπαίδευση και την αξιολόγηση του VGG Simple 1 μοντέλου. Ενδιαφέρον αποτελεί το γεγονός ότι το vgg_lstm_simple1 δεν καταφέρνει να αποδώσει ούτε αυτή τη φορά, καθώς ήταν το μόνο μοντέλο που δεν εμφάνισε σύγχλιση στην εκπαίδευση εγκυρότητας στην προηγούμενη ενότητα.

	precision	recall	f1-score	support
covid19	0.67	0.90	0.77	80
normal	0.93	0.94	0.93	905
pneumonia	0.95	0.89	0.92	800
accuracy			0.93	1785
macro avg	0.85	0.91	0.87	1785
weighted avg	0.92	0.92	0.92	1785

Table 4.24: vgg_simple 1

Στην συγκεκριμένη περίπτωση που ένα λανθασμένα αρνητικό τεστ έχει την μεγαλύτερη επίπτωση, το 90% της μετρικής recall ενθαρρύνει τη χρήση αυτού του μοντέλου. Όπως φαίνεται και από τους πίνακες η επιλογή της μεγαλύτερης ανάλυσης σε μια τέτοια περίπτωση οδηγεί σε καλύτερα αποτελέσματα. Οφείλει να σημειωθεί πως σε ένα άλλο σετ δεδομένων που η ενίσχυση της ποσότητας θα ήταν πολύ μεγαλύτερη (πχ 20000) εικόνες, το πιθανότερο θα ήταν να προτιμηθεί η ποσότητα, όπως επίσης και για δεδομένα τα οποία κάθε κλάση μπορεί να έχει τελείως διαφορετική κατανομή πχ κλάση "τοπία την ημέρα" - κλάση "τοπία τη νύχτα", τότε πάλι η ποσότητα θα ήταν η καλύτερη επιλογή. Στη συγκεκριμένη περίπτωση οι διαφορές των εικόνων σε κάθε κλάση είναι αρκετά μικρές, γεγονός που καθιστά την λεπτομέρεια στα χαρακτηριστικά απαραίτητη.

Κεφάλαιο 5

Σύνοψη και Προτάσεις

Στη παρούσα διπλωματική εργασία ερευνήθηκε η συμπεριφορά των μοντέλων βαθιάς μάθησης σε ελλιπή δεδομένα. Όπως είναι κατανοητό, ένα νέο φαινόμενο που προκύπτει και απασχολεί την ανθρωπότητα δεν θα διαθέτει πληθώρα πληροφοριών, τουλάχιστον στις αρχές του. Για αυτό το λόγο κάθε συνεισφορά και μεθοδολογία που υποδεικνύει η επιστημονική κοινότητα είναι πολύ σημαντική καθώς είτε θα είναι η αφετηρία για να κατανοηθεί πλήρως το φαινόμενο ή θα αποτελέσει τη βάση για να αντιμετωπιστεί ένα μελλοντικό παρεμφερές ζήτημα. Πιο συγκεκριμένα, είναι στατιστικά βέβαιο ότι ένας νέος ίός θα προκύψει σε μερικά χρόνια και ελπίζουμε να είναι αρκετά πιο ήπιος. Η παρούσα εργασία μπορεί να θεωρηθεί ως μια αναλυτική μεθοδολογία μελέτης και κατασκευής ενός μοντέλου ταξινόμησης ακτινογραφιών. Η συγκεκριμένη μελέτη πραγματοποιήθηκε με περιορισμένους πόρους τόσο σε αποθηκευτικό χώρο όσο και σε υπολογιστική δύναμη, δύο στοιχεία σχεδόν απαραίτητα στις σύγχρονες τεχνικές της βαθιάς μάθησης.

Συνοπτικά, η ροή της εργασίας ξεκίνησε από την αναζήτηση και συλλογή δεδομένων. Πρώτο βήμα ήταν η μελέτη σχετικών άρθρων για τον εντοπισμό των πιο δημοφιλών συνόλων δεδομένων με ακτινογραφίες θώρακα και ειδικά με ακτινογραφίες από ασθενείς με COVID-19. Αμέσως δημιουργήθηκε ο προβληματισμός από το μικρό πλήθος εικόνων COVID-19 όσο αφορά στην ικανότητα εκπαίδευσης ενός συστήματος βαθιάς μάθησης. Επιπλέον, δημιουργήθηκε η απορία ως προς το πώς ένα σύστημα που είναι ικανό να αφοριώσει τα χαρακτηριστικά των εικόνων, θα μπορούσε να είναι χρήσιμο όταν αυξηθούν κατά πολύ τα δεδομένα.

Μετά την συλλογή των διαθέσιμων συνόλων δεδομένων, εξετάστηκαν οι τρόποι με τους οποίους θα μπορούσε να αυξηθεί το πλήθος τους. Φυσικά η Επαύξηση Δεδομένων θα βοηθούσε αλλά θεωρήθηκε αρκετά ενδιαφέροντα η ιδέα δημιουργίας και χρήσης ενός Παραγωγικού Αντιπαλικού Δικτύου όχι μόνο για την παραγωγή "πλασματικών" εικόνων COVID-19 αλλά να είναι Υπό Συνθήκη (Conditional Gan), δηλαδή να μπορέσει να κατασκευάζει εικόνες από όλες τις κλάσεις, ανάλογα με τον κωδικό εισόδου του. Το κίνητρο προήλθε από το γεγονός ότι οπτικά όλες οι εικόνες του διαθέσιμου σετ μοιάζουν, κάτι που εκτιμήθηκε ως αδυναμία, αφού ένα ΠΑΔ αν εκπαιδευτεί μονάχα στην κλάση covid19 δεν θα έχει συναντήσει τις άλλες κλάσεις για να κατανοήσει τις διαφορές τους και να παράξει διαφορετικά αποτελέσματα. Κατασκευάστηκε επομένως το εν λόγω σύστημα,

ακολουθώντας τη δομή και το σκεπτικό του Υπό Συνθήκη ΠΑΔ [52], παρόλα αυτά δεν επιτεύχθηκε ικανοποιητικό οπτικό αποτέλεσμα, ίσως λόγω της περιορισμένης υπολογιστικής ισχύς, και επανεξετάστηκε η αρχική ιδέα ενός ΠΑΔ το οποίο εκπαιδεύεται σε μια μόνο κλάση. Αφού ενισχύθηκε το σετ δεδομένων, ακολούθησε η επιλογή προεκπαίδευμένων μοντέλων.

Η επιλογή των προεκπαίδευμένων μοντέλων στηρίζεται στην πορεία του διαγωνισμού ImageNet. Στο συγκεκριμένο διαγωνισμό παρατηρήθηκε για πρώτη φορά η χρήση συνελικτικών δικτύων με αρκετά στρώματα βάθος (VGG). Το βάθος αυτό οδήγησε στο πρόβλημα του vanishing gradient, το οποίο αντιμετωπίστηκε αποτελεσματικά με τα Residual Blocks (ResNet) και στη συνέχεια με τα Dense Blocks (DenseNet). Επίσης κάθε ένα από αυτά τα μοντέλα χαρακτηρίζεται από διαφορετική κλίμακα πλήθους μεταβλητών, το οποίο είναι αρκετά ενδιαφέρον για να εξεταστεί η αντίστοιχη συμπεριφορά τους στην αντιμετώπιση των συγκεκριμένων δεδομένων.

Αφού επιλέχθηκαν τα μοντέλα έπερπετε να βρεθεί τρόπος να επιλεχθεί μια καλή αρχιτεκτονική. Πρώτη σκέψη ήταν το k-fold cross validation[23] το οποίο και δοκιμάστηκε αλλά δυστυχώς ο χώρος αναζήτησης ήταν αρκετά μεγάλος και οι υπολογιστικοί πόροι δεν μπόρεσαν να το υποστηρίξουν. Επόμενη δοκιμή ήταν ο αλγόριθμος Βελτιστοποίησης Συμήνους Σωματιδίων ο οποίος δεν μπόρεσε να ολοκληρωθεί σε καμία από τις περιπτώσεις που δοκιμάστηκαν καθώς συνήθως εμφανιζόταν εξάντληση της μνήμης, αλλά μέχρι να συμβεί αυτό έδινε μια πολύ καλή εικόνα για το ποιος είναι ο πιθανός στόχος ακρίβειας για κάθε μοντέλο. Έτσι ορίστηκε ένα κριτήριο 85% ακρίβειας και κάθε φορά που η ΒΣΣ σχημάτιζε μοντέλο που ξεπερνούσε αυτόν τον αριθμό αποικεύοταν το μοντέλο. Όταν σταματούσε ο αλγόριθμος, εντοπίζοταν στα αποικεύμενα μοντέλα το βέλτιστο.

Αφού ολοκληρώθηκε αυτή η διαδικασία, αποφασίστηκε να επαναληφθούν τα βέλτιστα μοντέλα για να ελεγχθεί ότι ο αλγόριθμος ΒΣΣ είχε λειτουργήσει ορθώς. Παρατηρήθηκε ότι ορισμένα μοντέλα δεν μπόρεσαν να ξεπεράσουν σύτε το 45% στην ακρίβεια και χαρακτηριστικό ήταν το φαινόμενο να μην υπάρχει και αναβάθμιση των βαρών. Συνεπώς συλλέχθηκαν τα μοντέλα που μπόρεσαν να αποδώσουν εξίσου καλά και ολοκληρώθηκε το τελικό τεστ.

Στο τελικό τεστ το μοντέλο με τις περισσότερες μεταβλητές, το VGG19, κατάφερε να σημειώσει μεγαλύτερη ακρίβεια σε όλες τις πιθανές αρχιτεκτονικές συγκριτικά με τα υπόλοιπα προεκπαίδευμένα μοντέλα. Επίσης παρατηρήθηκε ότι το ενισχυμένο σύνολο δεδομένων δεν κατάφερε να αυξήσει την απόδοση των μοντέλων. Αυτό που διαπιστώθηκε στις περιπτώσεις των μοντέλων που εκπαιδεύτηκαν με το ενισχυμένο σύνολο δεδομένων είναι ότι περισσότερες δομές και αρχιτεκτονικές όχι έφταναν σε πιο υψηλή ακρίβεια συγκριτικά με τις λιγοστές περιπτώσεις των μοντέλων που εκπαιδεύτηκαν με το απλό σύνολο δεδομένων, η βέλτιστη όμως απόδοση των δύο ήταν ισοδύναμη.

Όσο αυξάνονται τα δεδομένα εικόνων COVID-19 (αυτή τη στιγμή το Covidx Dataset που χρησιμοποιήθηκε έχει φτάσει τις 900+ εικόνες) τόσο μεγαλύτερη είναι η δυνατότητα για καλύτερη εκπαίδευση των συστημάτων. Τα συστήματα αυτά όχι μπορούσαν είτε να είναι cGAN τα οποία όχι μπορέσουν να δώσουν πολλές πληροφορίες για τις διαφορές στις κατανομές των εικόνων με COVID-19 και άλλων ασθενειών ή σε συστήματα εντοπισμού όπως το R-CNN, έτσι ώστε να μην προκύπτει απλά θετική ή αρνητική διάγνωση, αλλά και εντοπισμός της περιοχής μόλυνσης που φυσικά μπορεί να βοηθήσει στη γρήγορη αντιμετώπιση του ασθενή από το ιατρικό προσωπικό. Τέλος, όλη η κινητοποίηση του τομέα

της βαθιάς μάθησης και της υπολογιστικής όρασης θα έχει ως αποτέλεσμα την δημιουργία πολλών μελλοντικών προεκπαιδευμένων μοντέλων που θα μπορέσουν να φανούν χρήσιμα σε μελλοντικές δύσκολες καταστάσεις, όπως αυτή που βιώνει η ανθρωπότητα το 2020.

Ο κώδικας της εργασίας μπορεί να βρεθεί εδώ [\[24\]](#).

Βιβλιογραφία

- [1] <https://machinelearningmastery.com/impressive-applications-of-generative-adversarial-networks/>. [Online].
- [2] http://cs231n.stanford.edu/slides/2017/cs231n_2017_lecture13.pdf. [Online].
- [3] http://cs231n.stanford.edu/slides/2017/cs231n_2017_lecture5.pdf. [Online].
- [4] <https://medium.com/thecyphy/gans-what-and-where-b377672283c5>. [Online].
- [5] <https://developers.google.com/machine-learning/gan/problems>. [Online].
- [6] <https://www.inference.vc/instance-noise-a-trick-for-stabilising-gan-training/>. [Online].
- [7] <https://www.tensorflow.org/tutorials/generative/cyclegan>. [Online].
- [8] http://plato.acadiau.ca/courses/comp/dsilver/NIPS95_LTL-transfer.workshop.1995.html. [Online].
- [9] <https://blog.keras.io/building-powerful-image-classification-models-using-very-little-data.html>. [Online].
- [10] <http://www.image-net.org/>. [Online].
- [11] <https://www.kaggle.com/getting-started/149448>. [Online].
- [12] <https://github.com/ieee8023/covid-chestxray-dataset>. [Online].
- [13] <https://github.com/agchung/Figure1-COVID-chestxray-dataset>. [Online].

- [14] <https://github.com/agchung/Actualmed-COVID-chestxray-dataset>. [Online].
- [15] <https://www.kaggle.com/tawsifurrahman/covid19-radiography-database>. [Online].
- [16] <https://www.kaggle.com/c/rsna-pneumonia-detection-challenge>. [Online].
- [17] <https://www.pythontinformer.com/python-libraries/numpy/numpy-and-images/>. [Online].
- [18] <https://github.com/philorfa/Covid19Project/blob/master/GANs/Covid19only.ipynb>. [Online].
- [19] <https://github.com/brucechou1983/CheXNet-Keras>. [Online].
- [20] <https://cv-tricks.com/keras/understand-implement-resnets/>. [Online].
- [21] <https://towardsdatascience.com/review-densenet-image-classification-b6631a8ef803>. [Online].
- [22] <https://towardsdatascience.com/understanding-and-visualizing-densenets-7f688092391a>. [Online].
- [23] <https://machinelearningmastery.com/k-fold-cross-validation/>. [Online].
- [24] <https://github.com/philorfa/Covid19Project>. [Online].
- [25] Yannis Avrithis, Nicolas Tsapatsoulis, and Stefanos Kollias. “Broadcast news parsing using visual cues: A robust face detection approach”. In: *2000 IEEE International Conference on Multimedia and Expo. ICME2000. Proceedings. Latest Advances in the Fast Changing World of Multimedia (Cat. No. 00TH8532)*. Vol. 3. IEEE. 2000, pp. 1469–1472.
- [26] Xi Chen et al. “Infogan: Interpretable representation learning by information maximizing generative adversarial nets”. In: *Advances in neural information processing systems*. 2016, pp. 2172–2180.
- [27] Scott C Douglas and Jiutian Yu. “Why RELU Units Sometimes Die: Analysis of Single-Unit Error Backpropagation in Neural Networks”. In: *2018 52nd Asilomar Conference on Signals, Systems, and Computers*. IEEE. 2018, pp. 864–868.

- [28] A Estava, B Kuprel, RA Novoa, et al. “Dermatologist level classification of skin cancer with deep neural networks [J]”. In: *Nature* 542 (2017), p. 115.
- [29] Muhammad Farooq and Abdul Hafeez. “Covid-resnet: A deep learning framework for screening of covid19 from radiographs”. In: *arXiv preprint arXiv:2003.14395* (2020).
- [30] Dario Garcia-Gasulla et al. “On the behavior of convolutional nets for feature extraction”. In: *Journal of Artificial Intelligence Research* 61 (2018), pp. 563–592.
- [31] Birte Glimm et al. “Lower and Upper Bounds for SPARQL Queries over OWL Ontologies.” In: *AAAI*. Citeseer. 2015, pp. 109–115.
- [32] Tripti Goel et al. “OptCoNet: an optimized convolutional neural network for an automatic diagnosis of COVID-19”. In: *Applied Intelligence* (2020), pp. 1–16.
- [33] Ian Goodfellow et al. “Generative adversarial nets”. In: *Advances in neural information processing systems*. 2014, pp. 2672–2680.
- [34] Venu G Gudise and Ganesh K Venayagamoorthy. “Comparison of particle swarm optimization and backpropagation as training algorithms for neural networks”. In: *Proceedings of the 2003 IEEE Swarm Intelligence Symposium. SIS’03 (Cat. No. 03EX706)*. IEEE. 2003, pp. 110–117.
- [35] Seung Seog Han et al. “Deep neural networks show an equivalent and often superior performance to dermatologists in onychomycosis diagnosis: Automatic construction of onychomycosis datasets by region-based convolutional deep neural network”. In: *PloS one* 13.1 (2018), e0191493.
- [36] Kaiming He et al. “Deep residual learning for image recognition”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 770–778.
- [37] Gao Huang et al. “Densely connected convolutional networks”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 4700–4708.
- [38] Aayush Jaiswal et al. “Classification of the COVID-19 infected patients using DenseNet201 based deep transfer learning”. In: *Journal of Biomolecular Structure and Dynamics* (2020), pp. 1–8.

- [39] James Kennedy and Russell Eberhart. “Particle swarm optimization”. In: *Proceedings of ICNN'95-International Conference on Neural Networks*. Vol. 4. IEEE. 1995, pp. 1942–1948.
- [40] Ilianna Kollia, Andreas-Georgios Stafylopatis, and Stefanos Kollias. “Predicting Parkinson’s disease using latent information extracted from deep neural networks”. In: *2019 International Joint Conference on Neural Networks (IJCNN)*. IEEE. 2019, pp. 1–8.
- [41] Ilianna Kollia et al. “Semantic image analysis using a symbolic neural architecture”. In: *Image Analysis & Stereology* 29.3 (2010), pp. 159–172.
- [42] Dimitrios Kollias, Athanasios Tagaris, and Andreas Stafylopatis. “On line emotion detection using retrainable deep neural networks”. In: *2016 IEEE Symposium Series on Computational Intelligence (SSCI)*. IEEE. 2016, pp. 1–8.
- [43] Dimitrios Kollias and Stefanos Zafeiriou. “Training deep neural networks with different datasets in-the-wild: The emotion recognition paradigm”. In: *2018 International Joint Conference on Neural Networks (IJCNN)*. IEEE. 2018, pp. 1–8.
- [44] Dimitrios Kollias and Stefanos P Zafeiriou. “Exploiting multi-cnn features in cnn-rnn based dimensional emotion recognition on the omg in-the-wild dataset”. In: *IEEE Transactions on Affective Computing* (2020).
- [45] Dimitrios Kollias et al. “Adaptation and contextualization of deep neural network models”. In: *2017 IEEE symposium series on computational intelligence (SSCI)*. IEEE, pp. 1–8.
- [46] Dimitrios Kollias et al. “Deep neural architectures for prediction in healthcare”. In: *Complex & Intelligent Systems* 4.2 (2018), pp. 119–131.
- [47] Dimitrios Kollias et al. “Photorealistic facial synthesis in the dimensional affect space”. In: *Proceedings of the European Conference on Computer Vision (ECCV)*. 2018, pp. 0–0.
- [48] Dimitris Kollias et al. “Interweaving deep learning and semantic techniques for emotion analysis in human-machine interaction”. In: *2015 10th International Workshop on Semantic and Social Media Adaptation and Personalization (SMAP)*. IEEE. 2015, pp. 1–6.

- [49] D Kollias et al. “Deep Transparent Prediction through Latent Representation Analysis”. In: *arXiv preprint arXiv:2009.07044* (2020).
- [50] Mohamed Loey, Florentin Smarandache, and Nour Eldeen M Khalifa. “Within the Lack of Chest COVID-19 X-ray Dataset: A Novel Detection Model Based on GAN and Deep Transfer Learning”. In: *Symmetry* 12.4 (2020), p. 651.
- [51] Thomas Lucas et al. “Mixed batches and symmetric discriminators for GAN training”. In: *International Conference on Machine Learning*. 2018, pp. 2844–2853.
- [52] Mehdi Mirza and Simon Osindero. “Conditional generative adversarial nets”. In: *arXiv preprint arXiv:1411.1784* (2014).
- [53] Jesús Muñoz-Bulnes et al. “Deep fully convolutional networks with random data augmentation for enhanced generalization in road detection”. In: *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE. 2017, pp. 366–371.
- [54] Saban Ozturk, Umut Ozkaya, and Mucahid Barstugan. “Classification of coronavirus images using shrunken features”. In: *medRxiv* (2020).
- [55] Hassan Mohamed Rabie. “Particle Swarm Optimization and Grey Wolf Optimizer to Solve Continuous p-Median Location Problems”. In: *International Conference on Advanced Intelligent Systems and Informatics*. Springer. 2019, pp. 136–146.
- [56] Alec Radford, Luke Metz, and Soumith Chintala. “Unsupervised representation learning with deep convolutional generative adversarial networks”. In: *arXiv preprint arXiv:1511.06434* (2015).
- [57] Konstantinos Rapantzikos et al. “Bottom-up spatiotemporal visual attention model for video analysis”. In: *IET Image Processing* 1.2 (2007), pp. 237–248.
- [58] Tim Salimans et al. “Improved techniques for training gans”. In: *Advances in neural information processing systems*. 2016, pp. 2234–2242.
- [59] Karen Simonyan and Andrew Zisserman. “Very deep convolutional networks for large-scale image recognition”. In: *arXiv preprint arXiv:1409.1556* (2014).

- [60] Abhijit Suresh, KV Harish, and N Radhika. “Particle swarm optimization over back propagation neural network for length of stay prediction”. In: *Procedia computer science* 46 (2015), pp. 268–275.
- [61] Arie Rachmad Syulistyo et al. “Particle swarm optimization (PSO) for training optimization on convolutional neural network (CNN)”. In: *Jurnal Ilmu Komputer dan Informasi* 9.1 (2016), pp. 52–58.
- [62] Athanasios Tagaris, Dimitrios Kollias, and Andreas Stafylopatis. “Assessment of Parkinson’s disease based on deep neural networks”. In: *International Conference on Engineering Applications of Neural Networks*. Springer. 2017, pp. 391–403.
- [63] Athanasios Tagaris et al. “Machine learning for neurodegenerative disorder diagnosis—survey of practices and launch of benchmark dataset”. In: *International Journal on Artificial Intelligence Tools* 27.03 (2018), p. 1850011.
- [64] Muhammed Talo et al. “Convolutional neural networks for multi-class brain disease detection using MRI images”. In: *Computerized Medical Imaging and Graphics* 78 (2019), p. 101673.
- [65] Linda Wang and Alexander Wong. “COVID-Net: A Tailored Deep Convolutional Neural Network Design for Detection of COVID-19 Cases from Chest X-Ray Images”. In: *arXiv preprint arXiv:2003.09871* (2020).
- [66] Ruoyao Wang et al. “High-resolution image reconstruction for portable ultrasound imaging devices”. In: *EURASIP Journal on Advances in Signal Processing* 2019.1 (2019), p. 56.
- [67] James Wingate et al. “Unified deep learning approach for prediction of Parkinson’s disease”. In: *IET Image Processing* 14.10 (2020), pp. 1980–1989.
- [68] Jun-Yan Zhu et al. “Unpaired image-to-image translation using cycle-consistent adversarial networks”. In: *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 2223–2232.